Medical Genetics

Volume I
Basic Genetics

Part I
Molecular Genetics

Dr. Mohammad Saad Zaghloul Salem
Professor Of Medical Genetics
Faculty Of Medicine, Ain-Shams University
Cairo, Egypt

2022

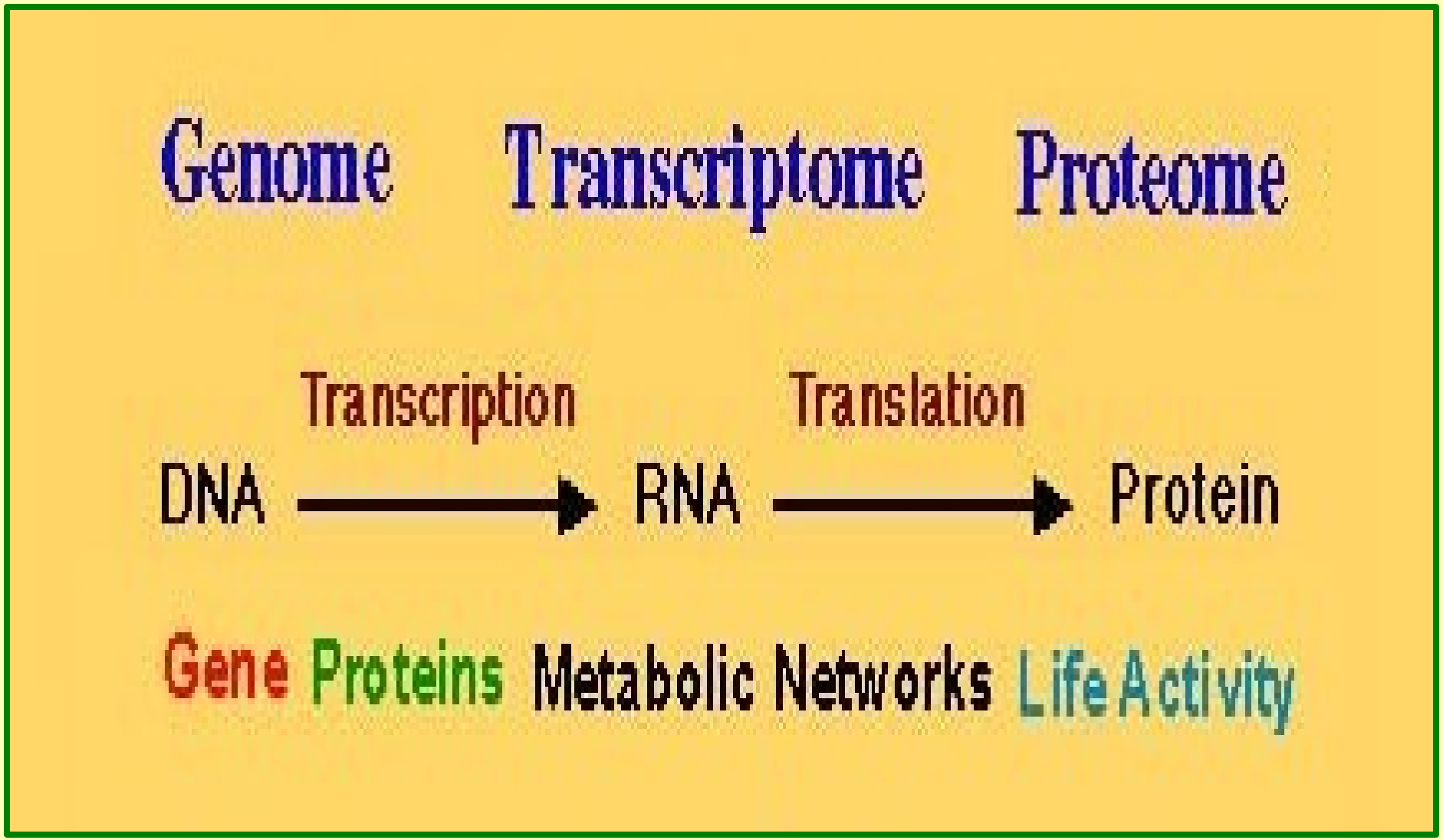
Spectrum of Medical Genetics

Basic Genetics	Clinical Genetics
Part I: Molecular Genetics Part II: Biochemical Genetics Part III: Physiological Genetics Part IIII: Cytogenetics Part V: Pathogenetics Part VI: Pharmacogenetics Part VII: Oncogenetics Part VIII: Immunogenetics Part IX: Formal Genetics Part X: Population genetics Part XI: Developmental Genetics Part XII: Genomics Part XIII: Transcriptomics Part XIV: Proteomics	Part I: Chromosomal Aberrations Part II: Congenital Malformations Part III: Inborn Errors of Metabolism Part IV: Mitochondrial Disorders Part V: Genetic Systemic Syndrome Part VI: Genetic Diseases of The Nervous system Part VII: Genetic Diseases of The Endocrinal system Part VIII: Genetic Diseases of The Cardio-Vascular system Part IX: Genetic Diseases of The Respiratory system Part X: Genetic Diseases of The Gastro-Intestinal system Part XI: Genetic Diseases of The Urinary system Part XII: Genetic Diseases of The Muscular system Part XIII: Genetic Diseases of The Skeletal system Part XIV: Genetic Diseases of The Blood system Part XV: Genetic Diseases of The Immunity system Part XVI: Genetic Diseases of The Male Genital system Part XVII: Genetic Diseases of The Female Genital system Part XVIII: Genetic Diseases of The Ocular system Part XIX: Genetic Diseases of The Auditory system Part XX: Genetic Diseases of The Skin Part XXI: Genetic Psychiatric Disorders
Diagnostic Genetics	**Therapeutic Genetics**
Part I: Molecular Diagnostic Techniques Part II: Cytogenetic Diagnostic Techniques Part III: Biochemical Diagnostic techniques Part IV: Prenatal Diagnosis Part V: Pre-Implantation Diagnosis Part VI: Per-Symptomatic Diagnosis Part VII: Conventional Diagnostic Techniques	Part I: Pharmacologic Therapy Part II: Nutritional Therapy Part III: Replacement Therapy Part IV: Transplantation Therapy Part V: Stem Cell Therapy Part VI: Surgical Intervention Part VII: Genetic Therapy Part VIII: Fetal Therapy Part IX: Conventional Therapy
Prophylactic Genetics	**Applied Genetics**
Part I: Pre-Conception Prophylaxis Part II: Pre-Natal Prophylaxis Part III: Pre-Symptomatic Prophylaxis	Part I: Forensic Genetics Part II: Genetic Counseling Part III: Genetic Screening Part IV: Genetic Engineering Part V: Eugenics

Dogma of Molecular Biology

Relation Between The genetic Material and Life Activities

Life Activities at The Molecular level



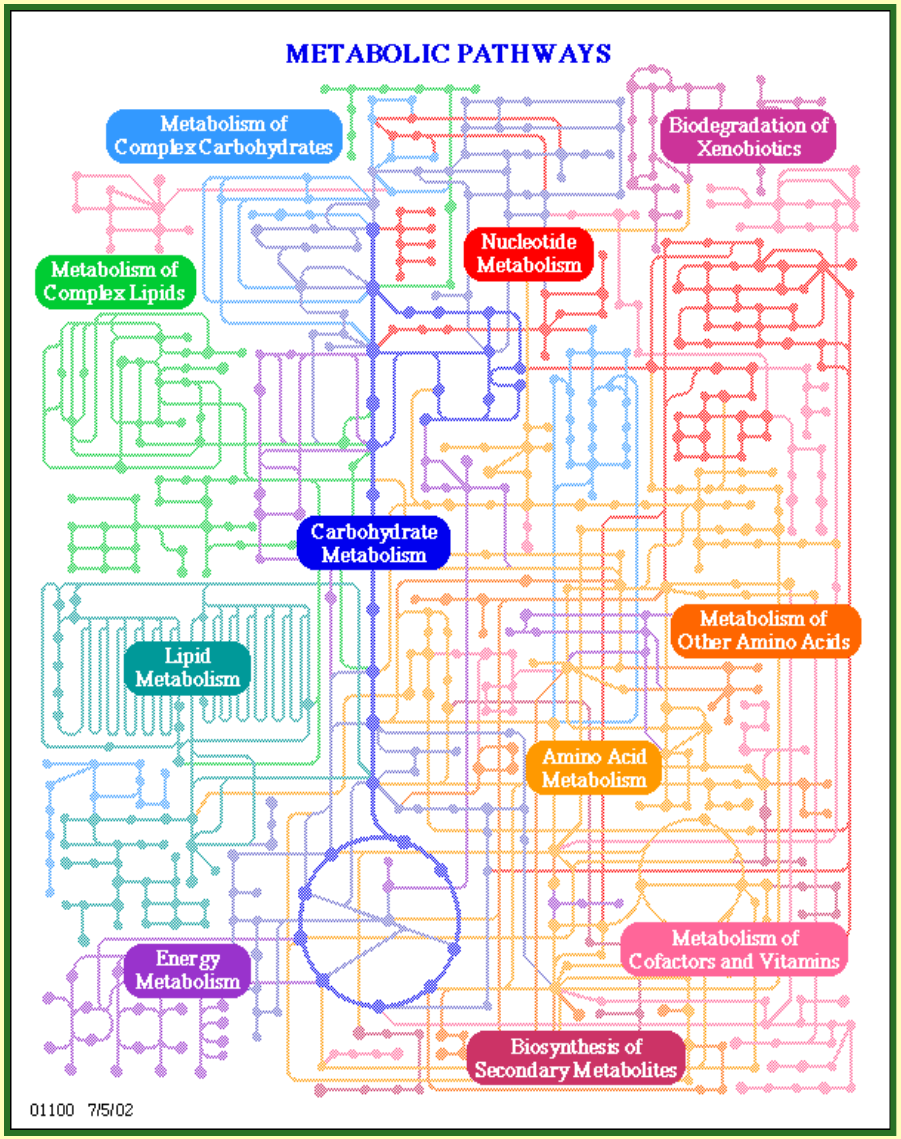
The Genetic Material And Life

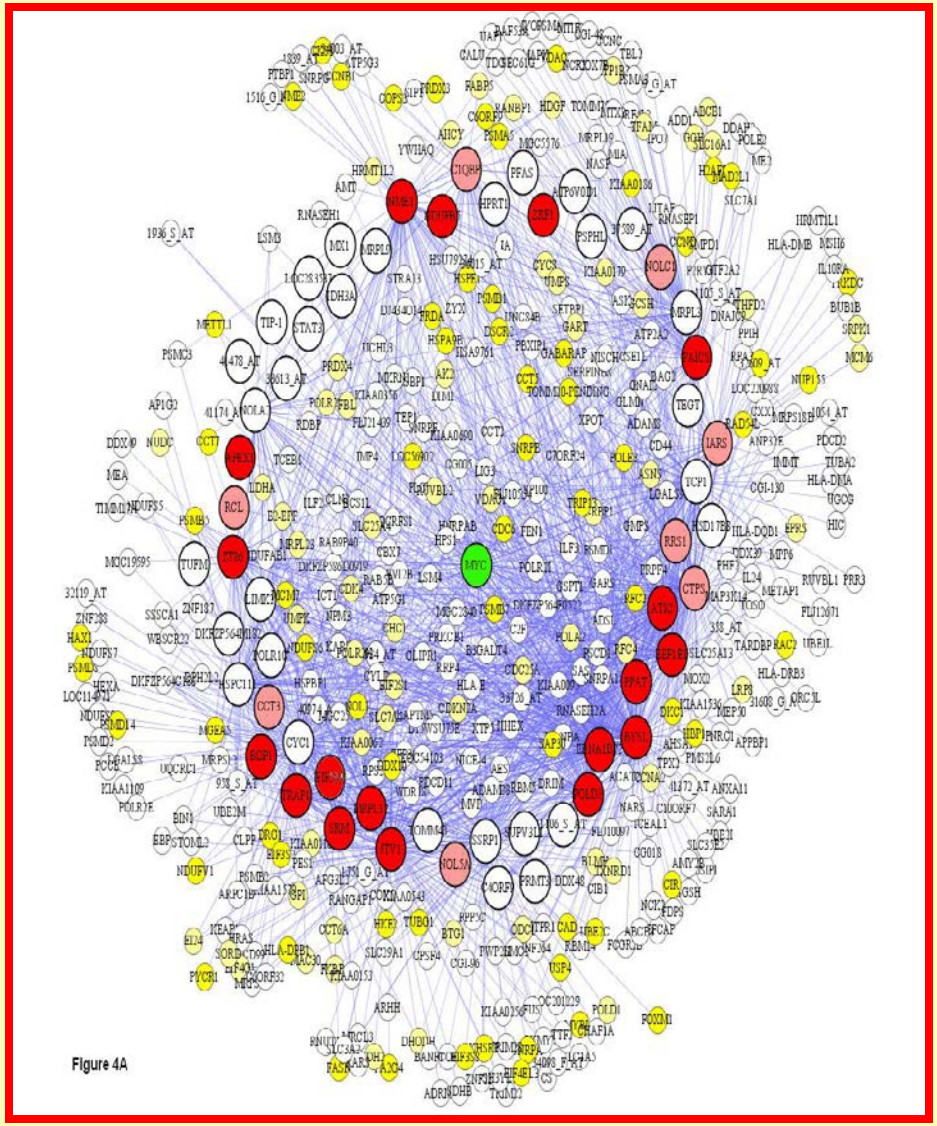
All life activities in living cells, whether on a molecular level like ATP production, a cellular level like cell division, a tissue level like muscle contraction or on a whole organ level like hearing for instance, are mediated via a very large number of inter-related metabolic networks. A metabolic network is defined as a cascade of controlled biochemical reactions and biophysical alterations that transform one, or more, substrate to one, or more, product. In human cells, nearly 4100 (four thousands and one hundred) of these networks have been delineated.

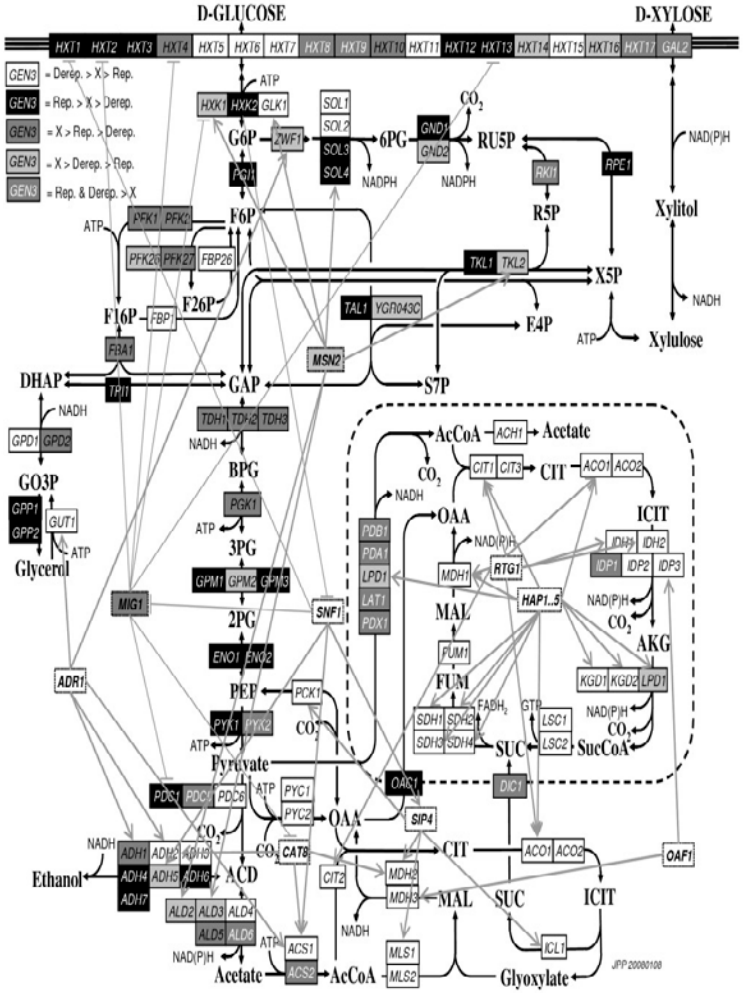
Each network consists of a very large number, sometimes thousands, of proteins, mostly enzymes, and other non-protein factors all acting co-operatively in sequence to perform specific biochemical and physiological functions.

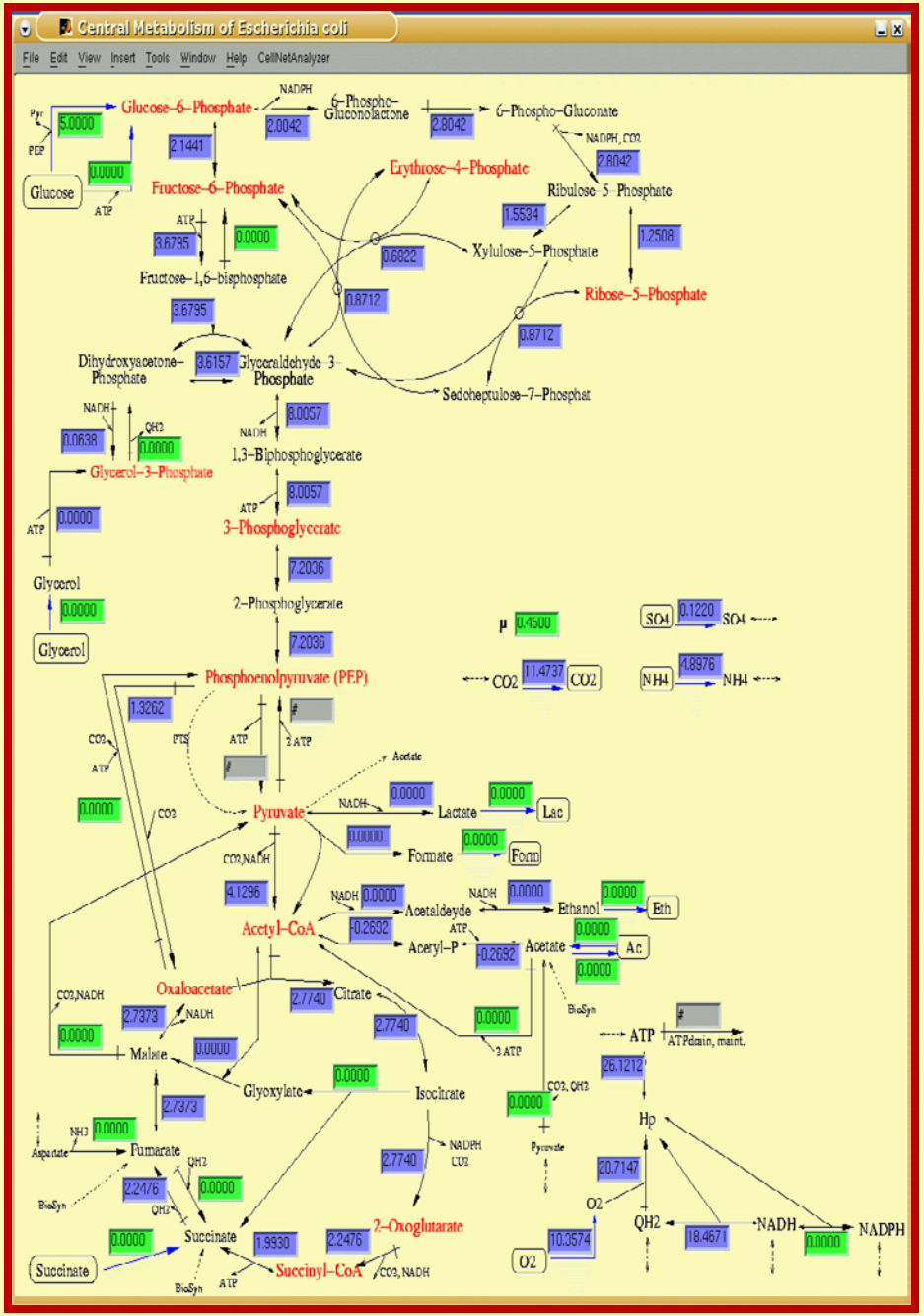
Proteins and enzymes which are the major mediators and determinants of all metabolic networks in living cells are synthesized under direct and strict regulation of the genetic material. The structural genes, which are the major component of the genetic material, are primarily concerned with controlling and regulating the synthesis of proteins, which in turn control and regulate life activities in cells.

The Concept Of Metabolic Networks

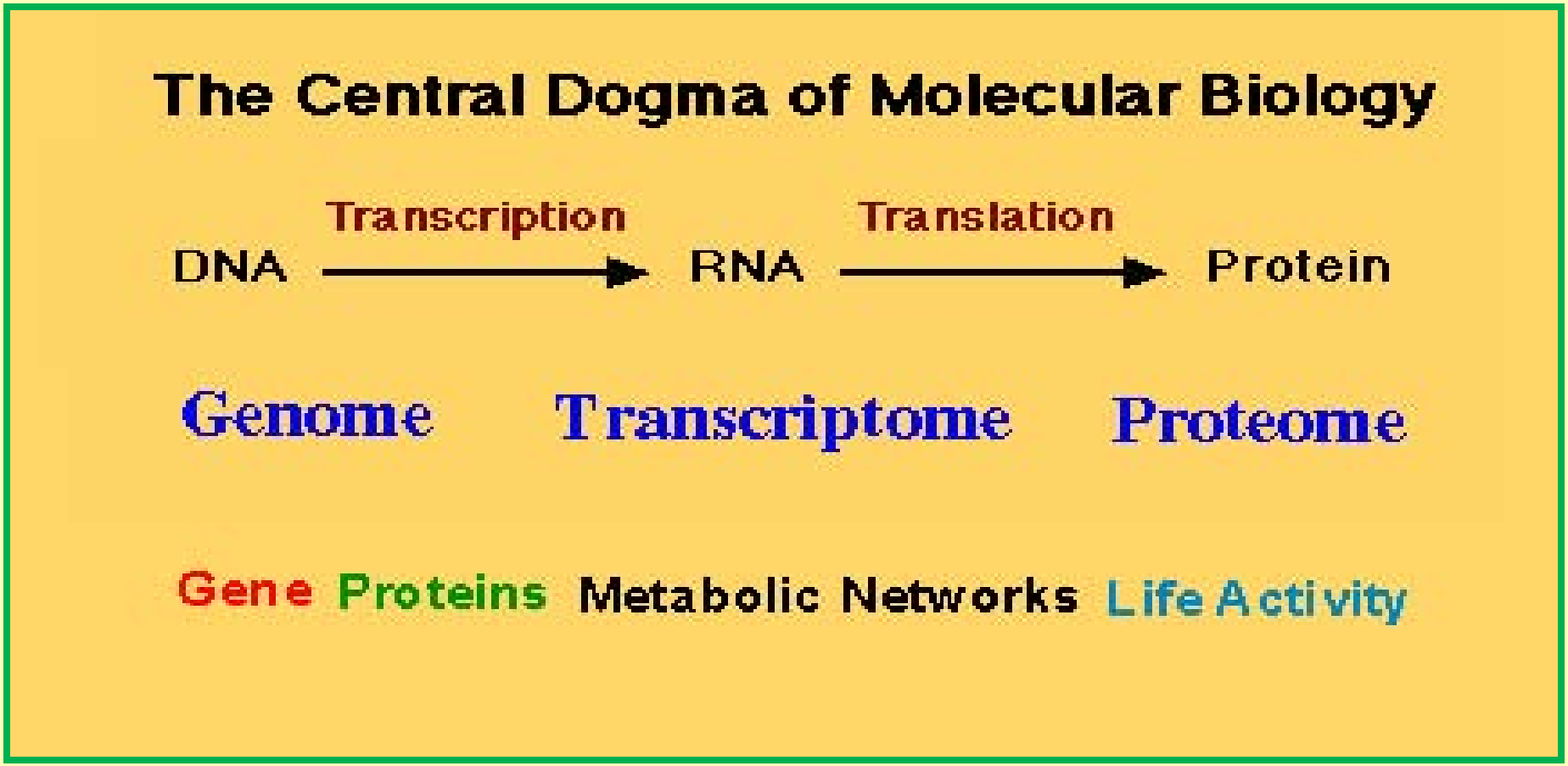





D-GLUCOSE
D-XYLOSE
HXT1 HXT2 HXT3 HXT4 HXT5 HXT6 HXT7 HXT8 HXT9 HXT10 HXT11 HXT12 HXT13 HXT14 HXT15 HXT16 HXT17 GAL2
GEN3 = Derep. > X > Rep.
GEN3 = Rep. > X > Derep.
GEN3 = X > Rep. > Derep.
GEN3 = X > Derep. > Rep.
GEN3 = Rep. & Derep. > X
ATP
HXK1 HXK2 GLK1
G6P
ZWF1
SOL1 SOL2 SOL3 SOL4
6PG
GND1 GND2
NADPH
CO2
RU5P
RKI1
RPE1
R5P
NAD(P)H
Xylitol
PGI1
F6P
PFK1 PFK2
PFK26 PFK27 FBP26
F26P
F16P
FBP1
FBA1
TKL1 TKL2
X5P
TAL1 YGR043C
F4P
NADH
Xylulose
MSN2
S7P
DHAP
TPI1
GAP
TDH1 TDH2 TDH3
GPD1 GPD2
GO3P
GPP1 GPP2
GUT1
Glycerol
BPG
PGK1
3PG
GPM1 GPM2 GPM3
2PG
ENO1 ENO2
PEP
PYK1 PYK2
Pyruvate
MIG1
SNF1
ADR1
PCK1
CO2
AcCoA
ACH1
Acetate
CIT1 CIT3
CIT
ACO1 ACO2
ICIT
PDB1
PDA1
LPD1
LAT1
PDX1
OAA
MDH1
RTG1
IDH1 IDH2
IDP1 IDP2 IDP3
HAP1..5
MAL
FUM1
FUM
KGD1 KGD2 LPD1
AKG
SDH1 SDH2 SDH3 SDH4
FADH2
GTP
LSC1 LSC2
SUC
SucCoA
OAC1
DIC1
PDC1 PDC5 PDC6
PYC1 PYC2
SIP4
OAF1
CAT8
CIT2
MDH2 MDH3
ADH1 ADH2 ADH3 ADH4 ADH5 ADH6 ADH7
ACD
ALD2 ALD3 ALD4 ALD5 ALD6
ACS1 ACS2
MLS1 MLS2
ICL1
JPP 20080108


Central Metabolism of Escherichia coli
File Edit View Insert Tools Window Help CellNetAnalyzer
Glucose-6-Phosphate
6-Phospho-Gluconolactone
6-Phospho-Gluconate
Erythrose-4-Phosphate
Ribulose-5-Phosphate
Fructose-6-Phosphate
Xylulose-5-Phosphate
Fructose-1,6-bisphosphate
Ribose-5-Phosphate
Dihydroxyacetone-Phosphate
Glyceraldehyde-3-Phosphate
Sedoheptulose-7-Phosphat
Glycerol-3-Phosphate
1,3-Biphosphoglycerate
3-Phosphoglycerate
Glycerol
2-Phosphoglycerate
Phosphoenolpyruvate (PEP)
Pyruvate
Lactate
Formate
Acetaldeyde
Ethanol
Acetyl-CoA
Acetate
Oxaloacetate
Citrate
Malate
Glyoxylate
Isocitrate
Fumarate
Succinate
2-Oxoglutarate
Succinyl-CoA
ATP
ATPdain, maint.
Hp
NADH
NADPH
O2
CO2
NH4
SO4

Hence, though the genetic material controls and encompass the whole spectrum of life processes in living cells, the proteins are the actual and direct mediators of these life processes.



Dogma Of Molecular Pathology In Health And Disease

Genome Transcriptome Proteome

DNA —Transcription→ RNA —Translation→ Protein

Gene Proteins Metabolic Networks Life Activity

Mutant Gene ⟹ Abnormal mRNA ⟹ Deficient/Defective Excess Product

Abnormal Metabolic Networks ⟹ Disturbed Cell Function ⟹

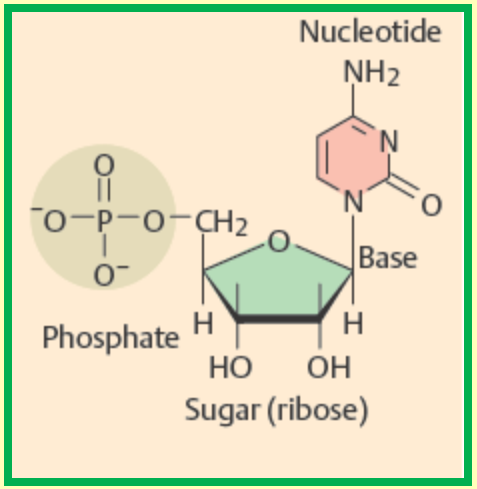
Deranged Physiological Activities ⟹ Disease ⟹ Cancer

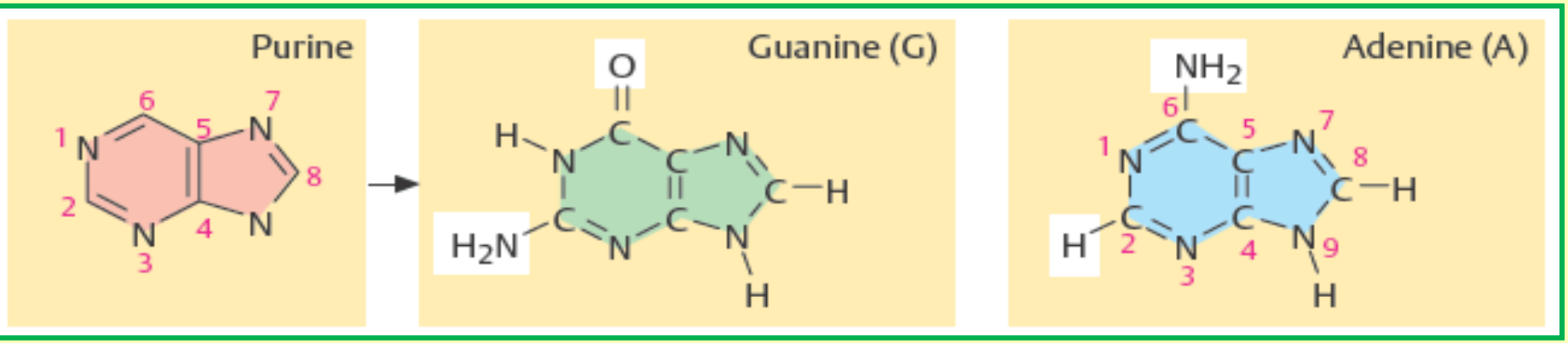
Genetic Disorder Congenital Anomaly Immunodeficiency

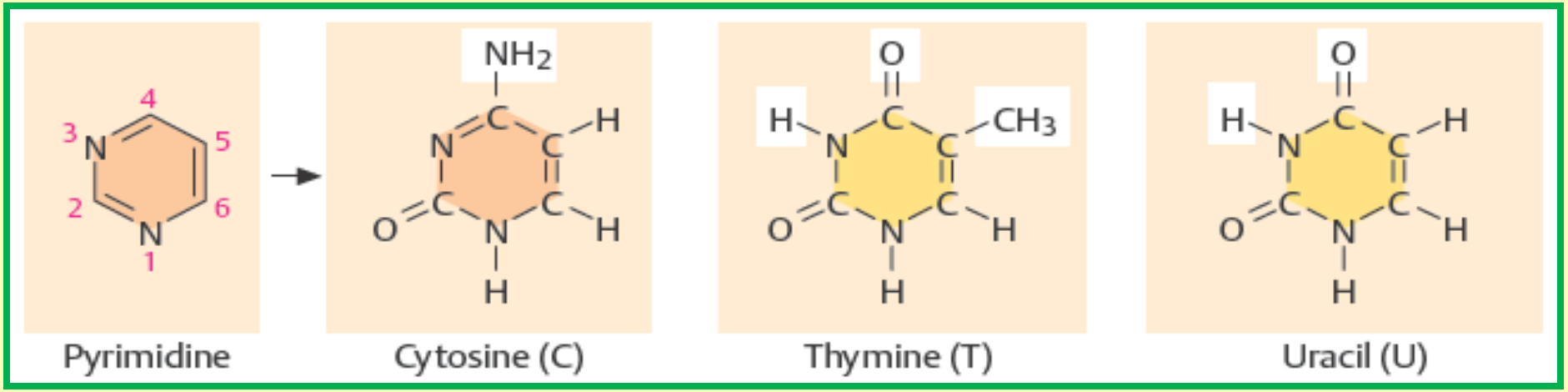
Structure Of The Genetic Material

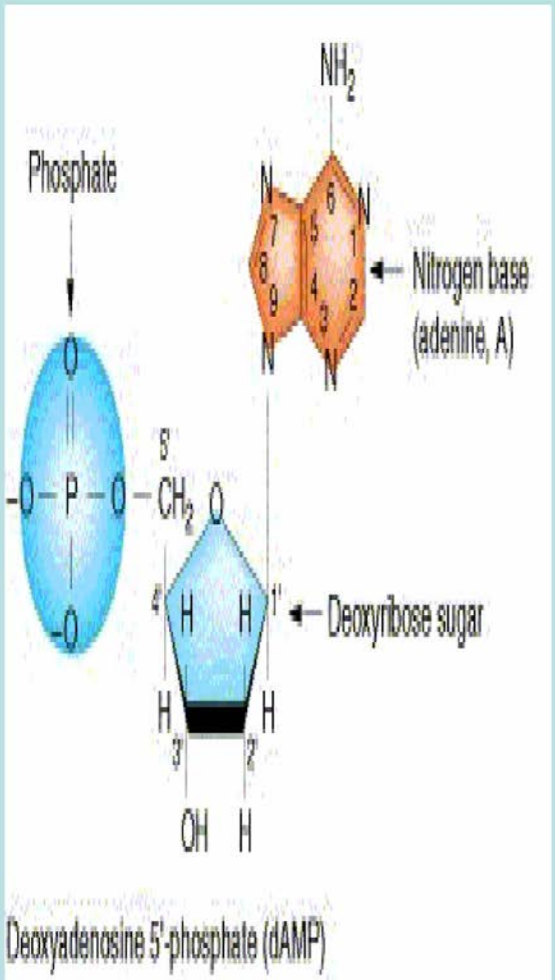
The building components of the genetic material in all living creatures are the nucleic acids. There are two main categories of nucleic acids : DNA or Deoxyribo-Nucleic Acid and RNA or Ribo-Nucleic Acid. With the exception of RNA-viruses which have their genome composed solely of RNA, all living creatures have DNA as their sole genetic material in addition to RNA as well.

Nucleic acids are very long unbranched hetero-polymers, composed of large number of similar monomers : the nucleotides, which are the building blocks of the nucleic acids.

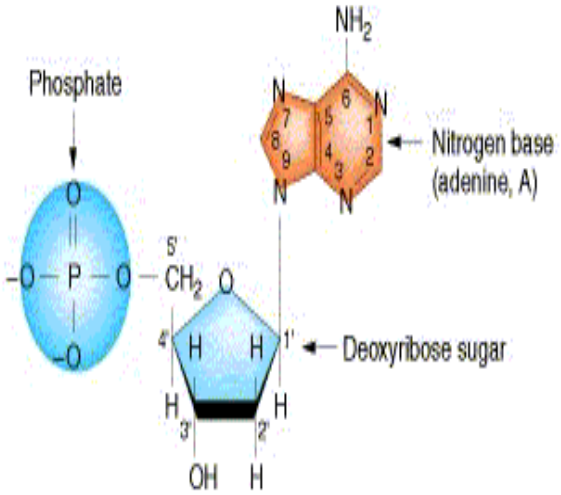

Nucleotide
Base
Phosphate
Sugar (ribose)


Purine
Guanine (G)
Adenine (A)

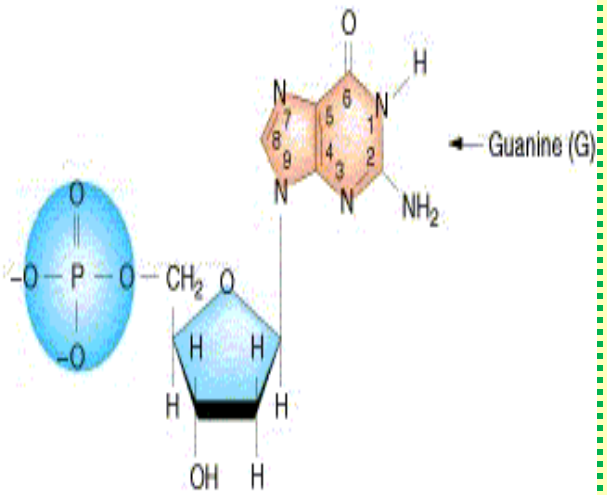

Pyrimidine
Cytosine (C)
Thymine (T)
Uracil (U)



Purine nucleotides

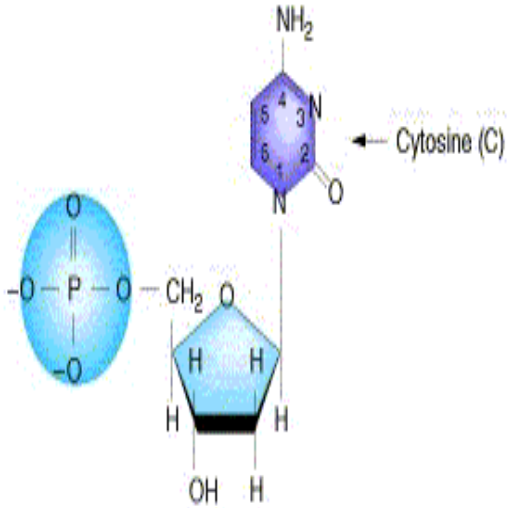


Deoxyadenosine 5'-phosphate (dAMP)

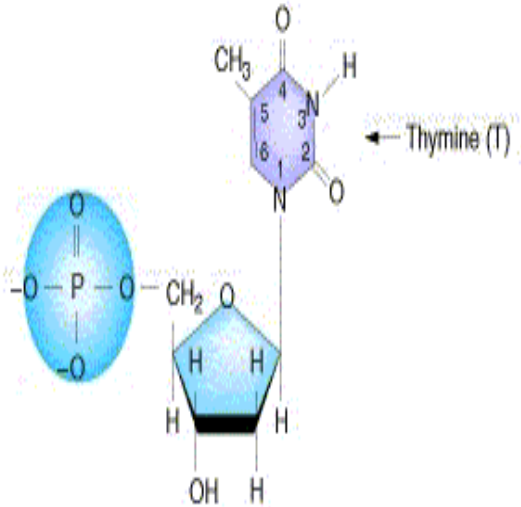


Deoxyguanosine 5'-phosphate (dGMP)

Pyrimidine nucleotides

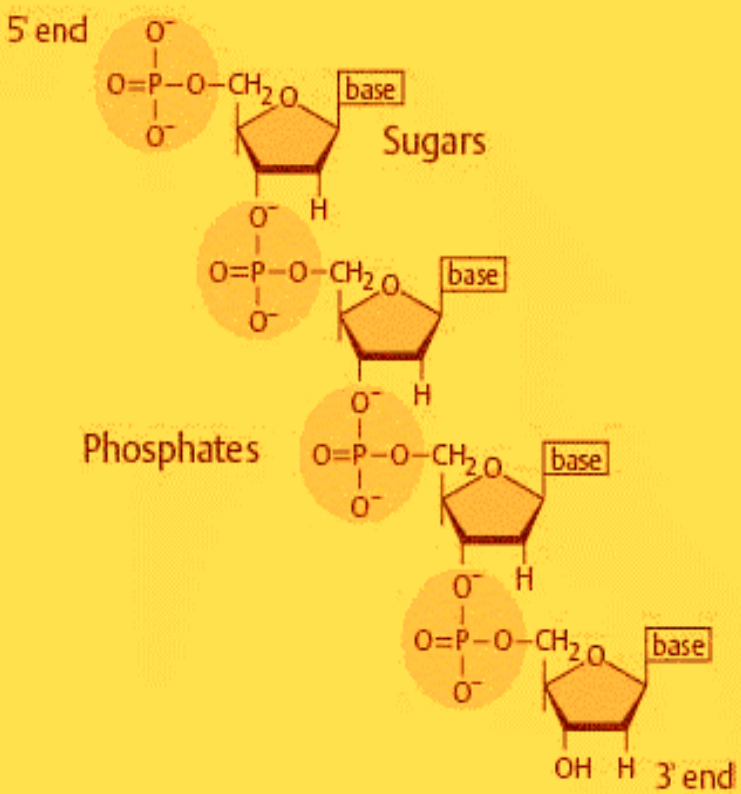


Deoxycytidine 5'-phosphate (dCMP)

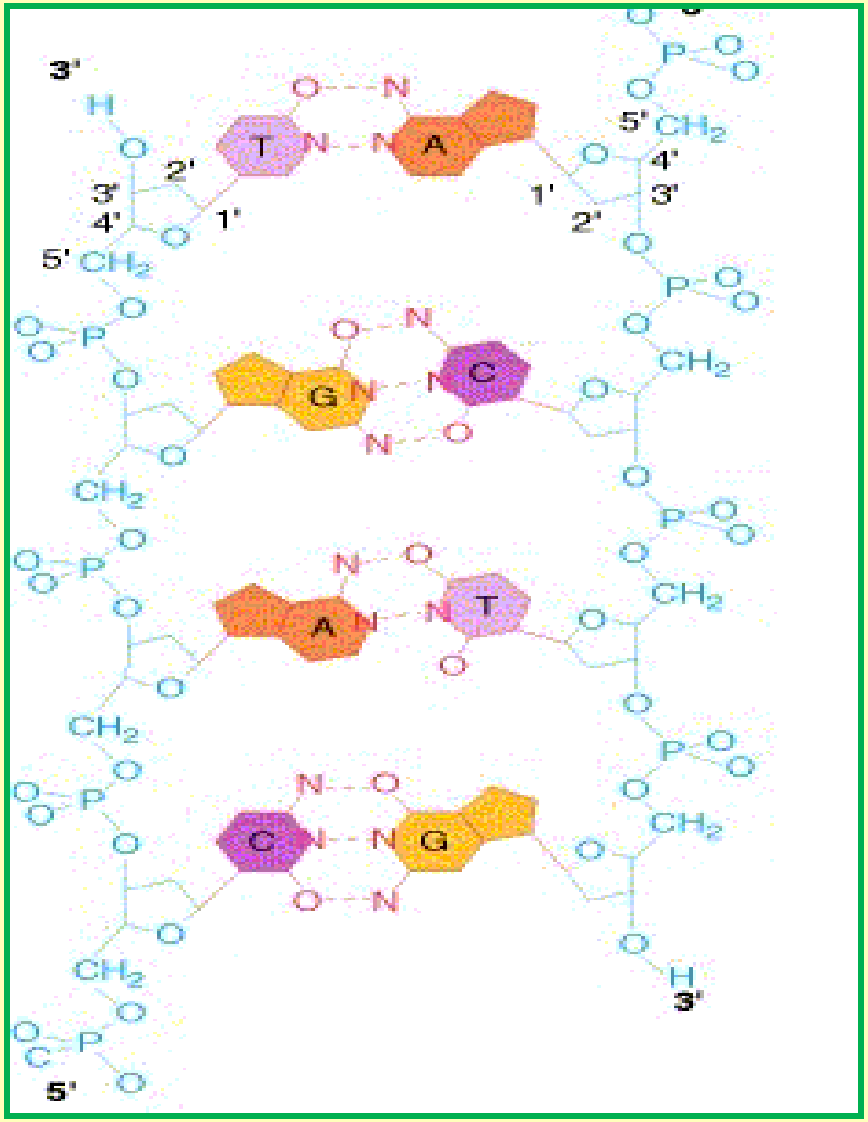


Deoxythymidine 5'-phosphate (dTMP)

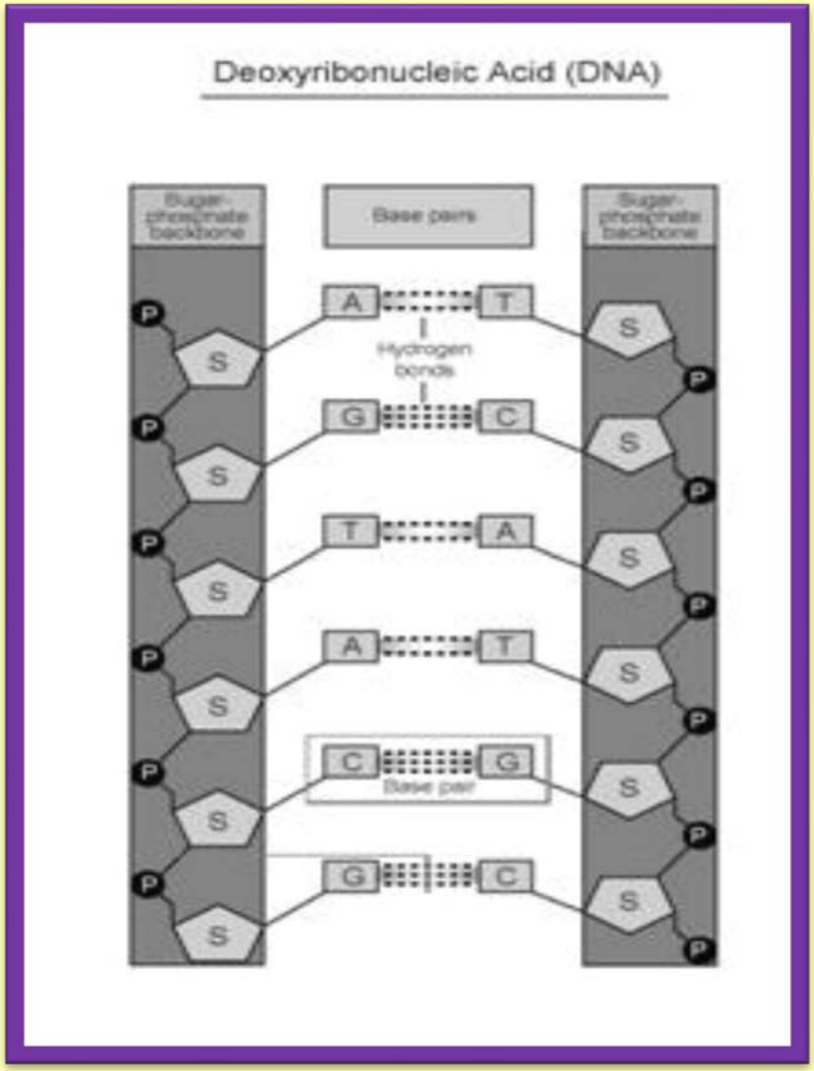
Structure Of The Genetic Material
The Nucleic Acids

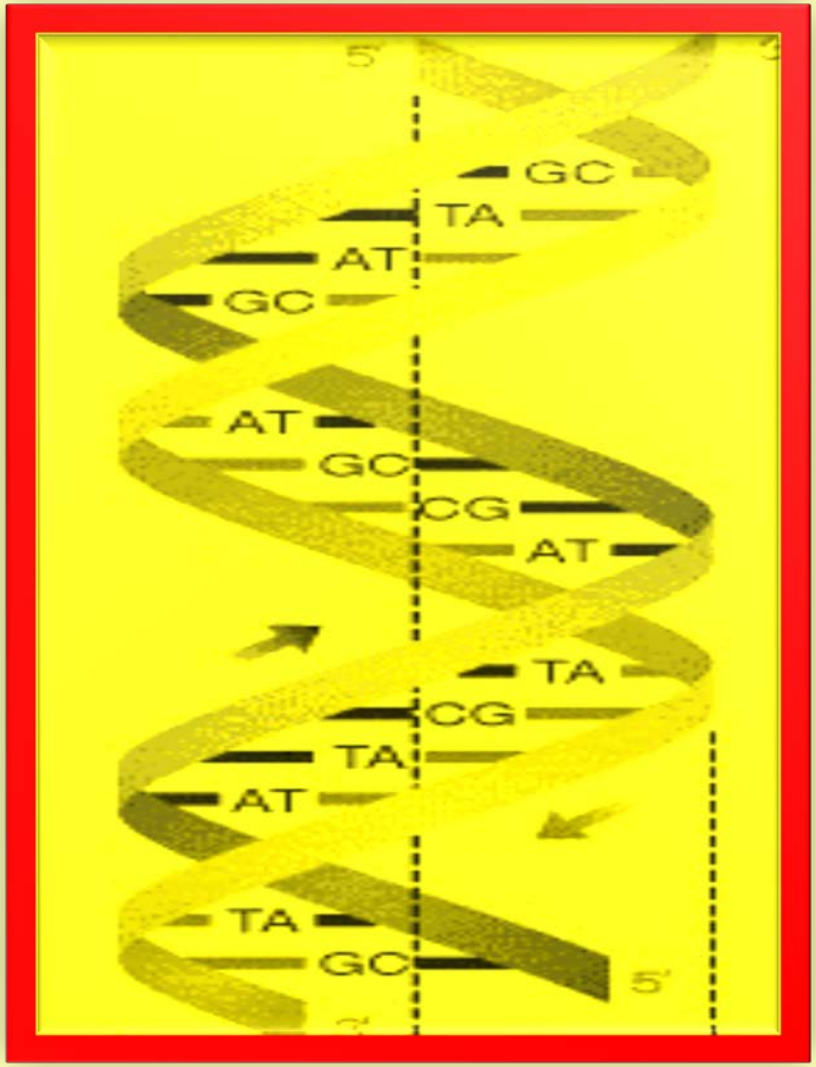


The longitudinal strand-shaped structure of the nucleic acids

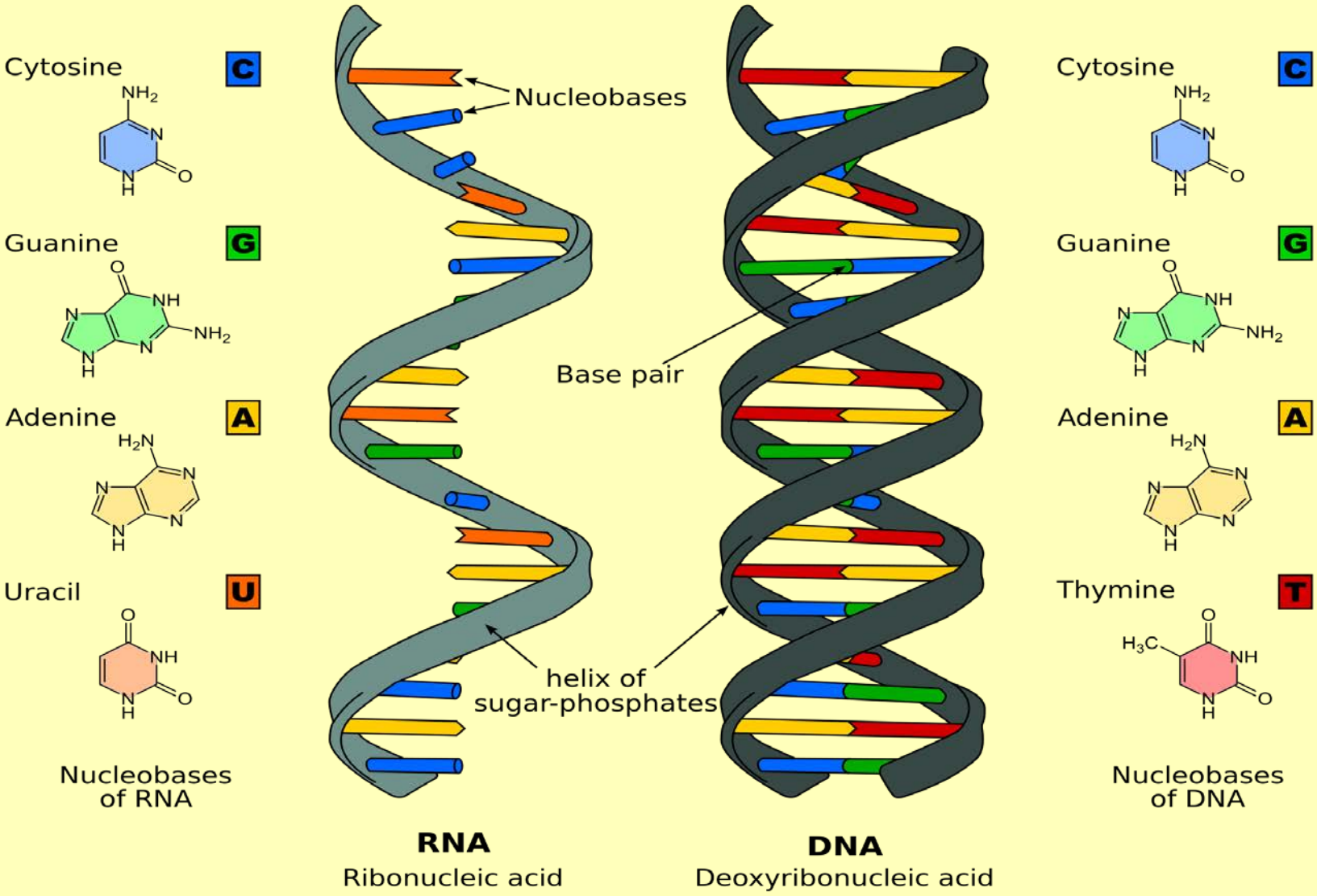


DNA Structure & The Concept Of Base Complementarity





Structure of Nucleic Acids



Each nucleotide is composed of an inorganic phosphate group attached to a 5-carbon atom sugar, ribose sugar in RNA and 2-deoxyribose sugar in DNA, to which is attached a nitrogenous base.

Five different bases participate in formation of five different nucleotides that build up the nucleic acids. The bases are either purine bases : adenine (A) and guanine (G), or pyrimidine bases : cytosine (C), thymine (T), and uracil (U). The nucleotides are usually referred to by the type of base they contain, hence we have (T), (C), (G), (A) and (U) nucleotides. The first four nucleotides are found exclusively in DNA, and Uracil replaces Thymine in RNA.

The longitudinal strand-shaped structure of the nucleic acids is maintained by the side-by-side attachment of the nucleotides, with the phosphate group of one nucleotide being attached to the ribose sugar of the next nucleotide.

DNA occurs naturally as a double stranded structure composed of two complementary strands attached together by the hydrogen bonds of the nitrogenous bases of each two opposing nucleotides. With few exceptions, RNA exists as a single stranded structure.

Structural Differences Between DNA & RNA

RNAs differ from DNAs in many aspects :

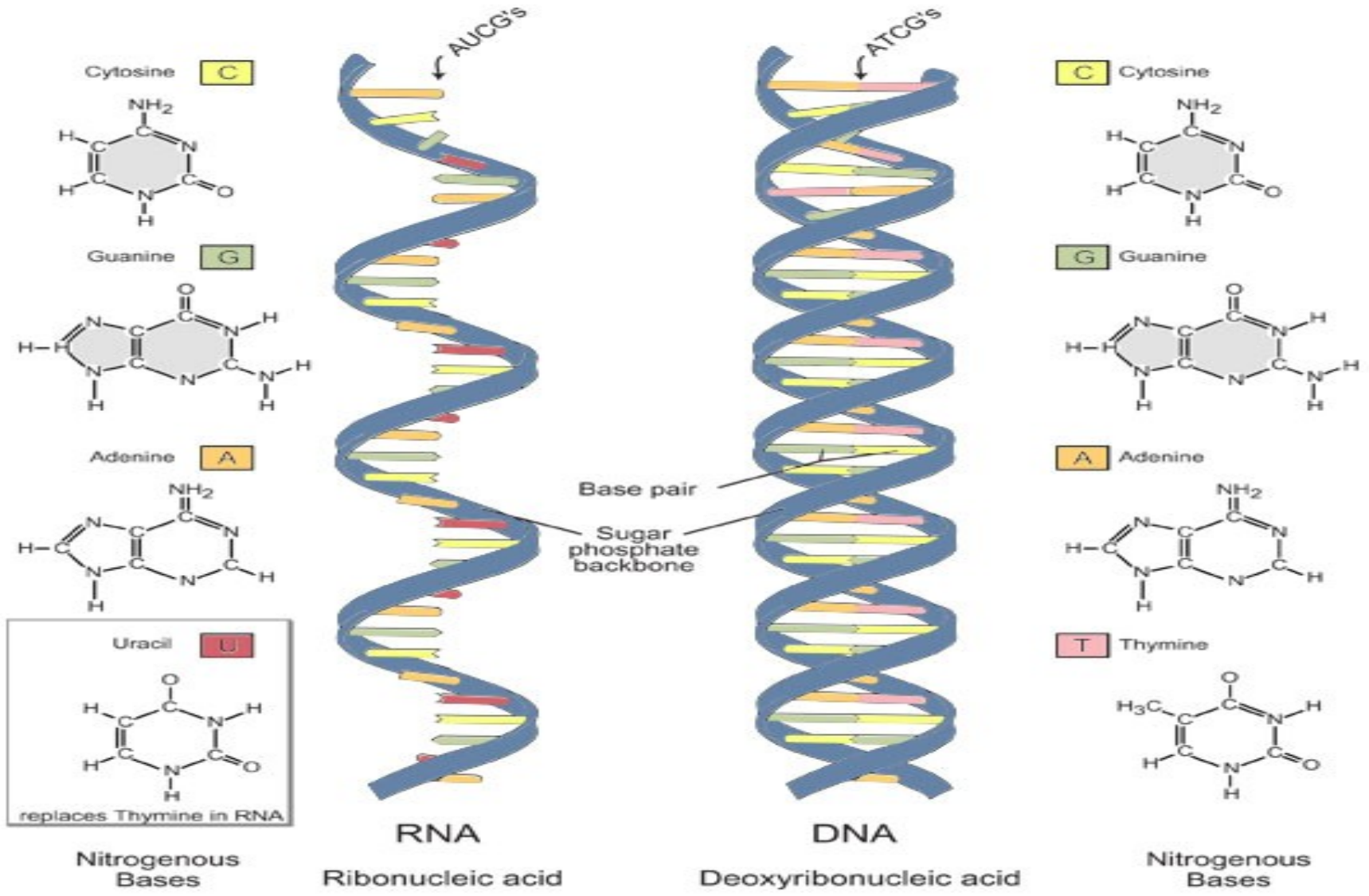
1- Most RNAs are single stranded molecules with the exception of some types of doubles stranded small or micro RNAs.

2- They have Uracil (U) instead of Thymine (T).

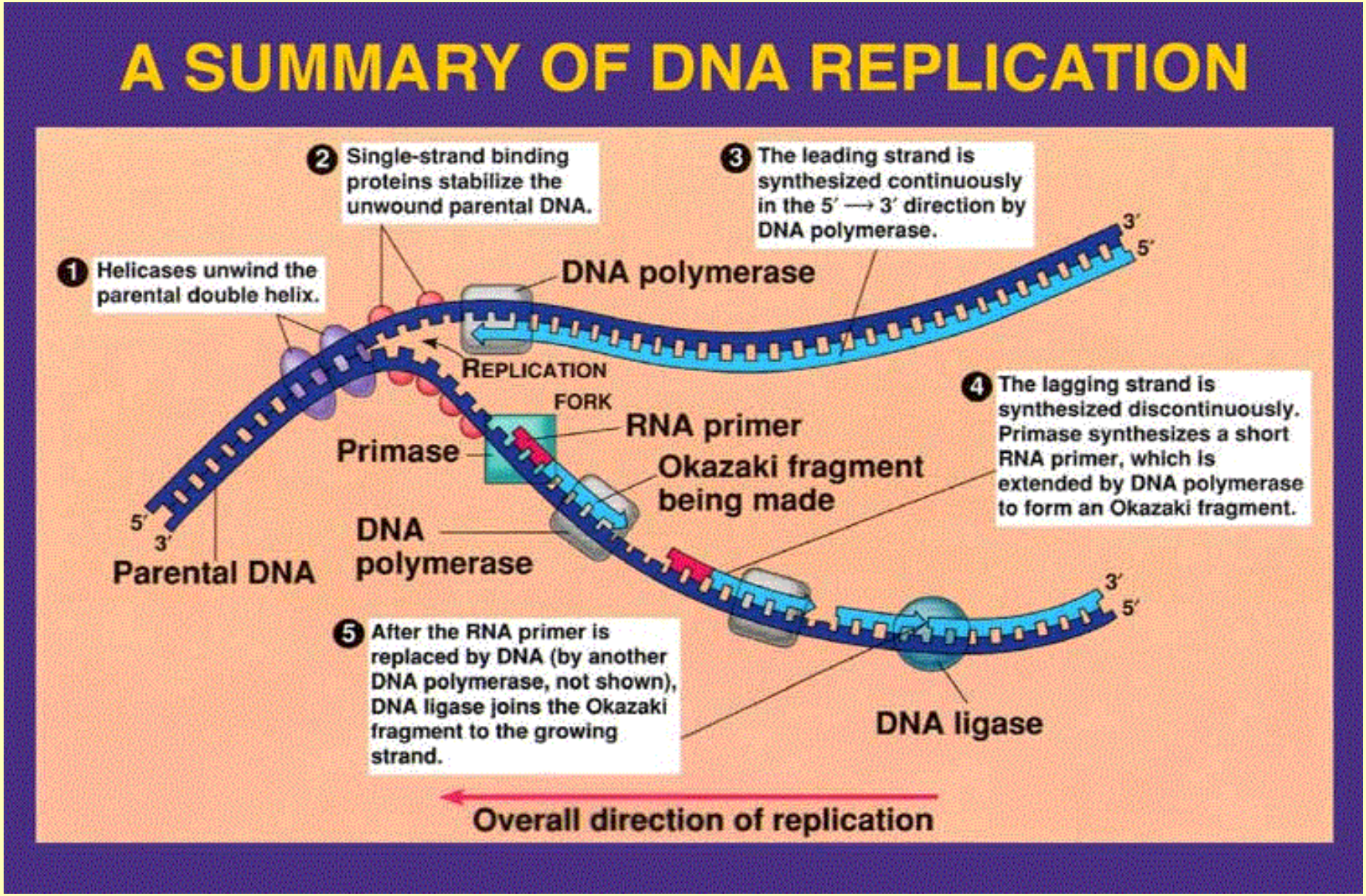
3- They have Ribose sugar instead of 2-Deoxyribose sugar.

4- There are many types of RNAs : messenger (mRNA), ribosomal (rRNA), transfer (tRNA) and small or micro RNA (miRNA). DNA exists as one type albeit with different structural configurations.

5. RNAs exist in the nucleus and the cytoplasm but DNA exists only within the cell nucleus and the mitochondria.


Cytosine C
Guanine G
Adenine A
Uracil U
replaces Thymine in RNA
Nitrogenous Bases
AUCG's
RNA
Ribonucleic acid
Base pair
Sugar phosphate backbone
ATCG's
DNA
Deoxyribonucleic acid
C Cytosine
G Guanine
A Adenine
T Thymine
Nitrogenous Bases

Mechanism Of DNA Replication



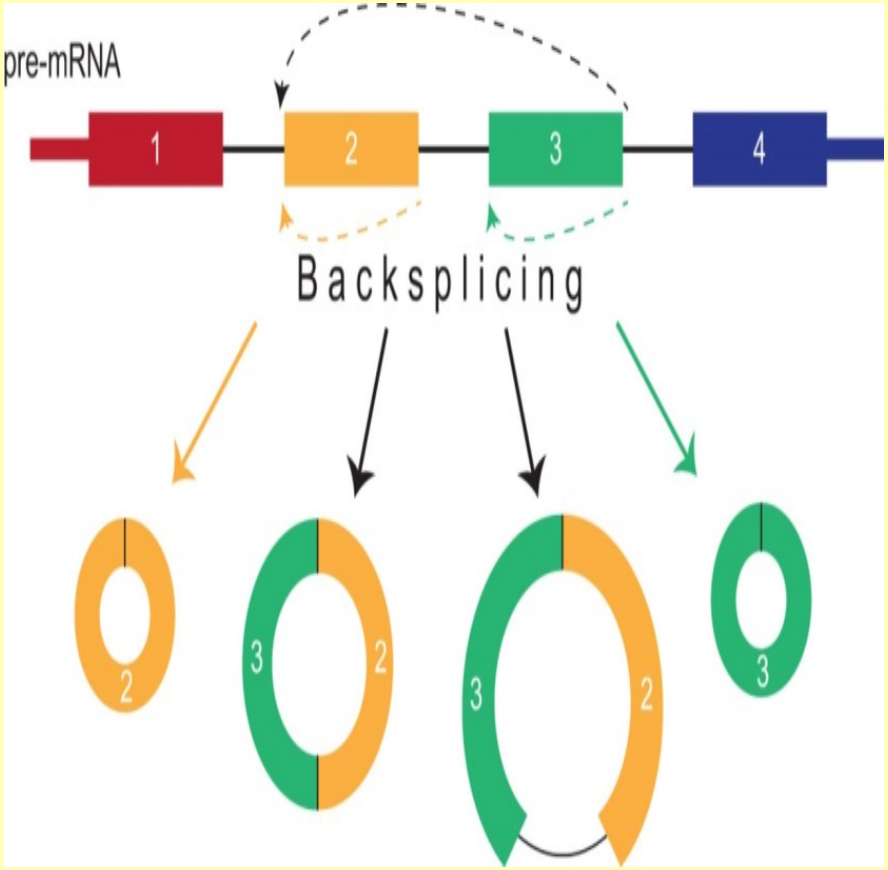
Functional categories of RNA

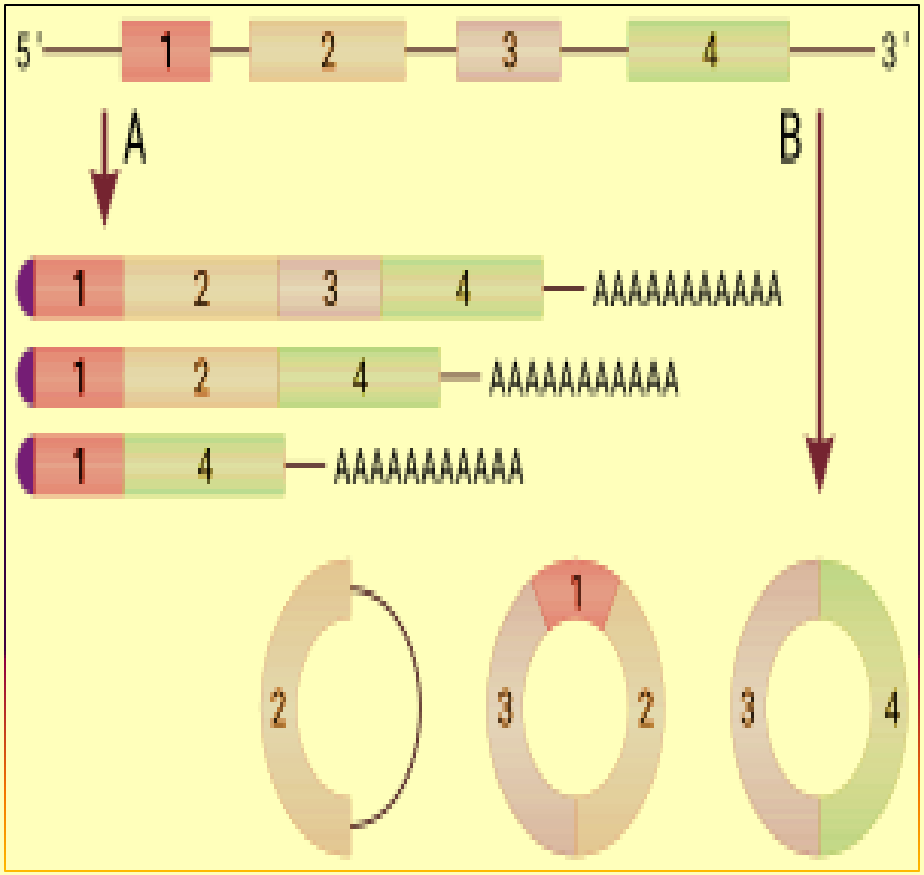
Currently, at least six main functional subtypes of RNA have been well characterized, both structurally and functionally. These subtypes include:

1. Messenger RNA (mRNA) which is the main product and mediator of transcription, carrying the information necessary for protein synthesis.
2. Ribosomal RNA (rRNA) which functions in translation via decoding the mRNA code to recognize the amino acid defined by the specific codon.
3. Transfer RNA (tRNA) which also functions in translation via decoding the mRNA code to recognize the amino acid defined by the specific codon in addition to getting the amino acid from the cytosol to site of protein synthesis.

4. Circular RNAs (circRNA) species probably fulfill diverse biological functions including regulation of transcription and modulation of protein-RNA binding. circRNAs usually result from splicing events, either as exonic circRNA from circularization of exons or as intronic circRNA, such as, for example, circular tRNA and circular rRNA introns produced from archaeal splicing. In vitro, RNA circularization involves the intra-molecular formation of a 3′, 5′-phosphodiester bond, requiring close proximity of the 3′- and 5′-terminus of the linear precursor. The circular form, rather than the traditional linear form, of circRNA confers marked stability to the molecule because it protects it from degradation by ubiquitously spread cytoplasmic exonucleases because of lack of the polyadenylate tail which is the primary target of degradation. The expression of circRNAs is developmentally regulated, tissue and cell-type specific, and shared across the eukaryotic tree of life. These features suggest important functions for these molecules. Also, the dynamic and cell type-specific expression patterns during development of circRNAs suggest potential developmental roles and functions of circRNAs, specially during brain and neuronal development.

Mechanisms of Circular RNAs Formation





5. Piwi-interacting RNAs (piRNA) constitute the largest class of small non-coding RNA species expressed in animal cells in both vertebrates and invertebrates, and it is estimated that mammalian cells contain many hundreds of thousands of different piRNA species. They are distinct from microRNA in many aspects, viz, larger size (26-31 rather than 21-24 nucleotides) with 5' monophosphate and peculiar 3' modification (2'-O-methylation) that has been suggested to increase stability of the molecule, probably via reducing its destruction by active oxidant radicals, lack of sequence conservation, increased structural complexity, and a biogenesis pathway clearly distinct from that of miRNA.

piRNAs form RNA-protein complexes through interactions with piwi proteins. These piRNA complexes have been linked to both epigenetic and post-transcriptional silencing of retrotransposons and other genetic elements in germ line cells, particularly those in spermatogenesis. They play fundamental roles in stabilizing the genome due to their roles in suppressing and silencing increased transposon activities during development, hence they probably exert a vital protective role against transposon-induced teratogenesis and development of congenital malformations.

piRNA comprises many subspecies found in the nucleus and the cytoplasm of germ cells, particularly in male germ cell lines., e.g., small interfering RNA (siRNA) which play critical roles in regulating gene expression and translation. In view of being transmitted maternally, they may be involved in maternally derived epigenetic effects.

piwi-interacting RNA (piRNA)

piwiRNA, are composed of RNA-piwi protein complexes and constitutes the largest portion of small non-coding RNA molecules in animal cells, in both vertebrates and invertebrates. There are many hundreds of thousands of different piwiRNA species in mammals. They exert important roles in epigenetic and post-transcriptional gene silencing of retro-transposons and other genetic elements in germ line cells, particularly those in spermatocytes during spermatogenesis.

piRNAs are composed of (26–31) nucleotides with a 5’ monophosphate and a 3’ modification (2’-O-methylation) that has been suggested to increase stability of the molecule, probably via reducing its destruction by oxidant radicals. piRNA classes do not have secondary structure, they lack sequence conservation and comprise many subtypes that are found in the nucleus and the cytoplasm of germ cells, particularly in male germ cell lines.

piRNAs are thought to be involved in gene silencing, specifically the silencing of transposons. The majority of piRNAs are antisense to transposon sequences suggesting that transposons are the main target of piRNA. In mammals it appears that the activity of piRNAs in transposon silencing is most important during the development of the embryo.

6. Micro RNA or Small RNA (miRNA) are non-coding species of RNA, i.e. they are not transcripts of genes and are not translated to proteins. Small or micro RNAs comprise many different subtypes that play critical roles in many functional aspects of the genetic material including regulation of transcription, regulation of post-transcription silencing/enhancement of translation, regulation of protein trafficking, and many other critical processes.

Subtypes of miRNAs include:

a. **Guide RNA (gRNA)** involved in RNA editing and correction of transcription errors involving some specific point mutations.

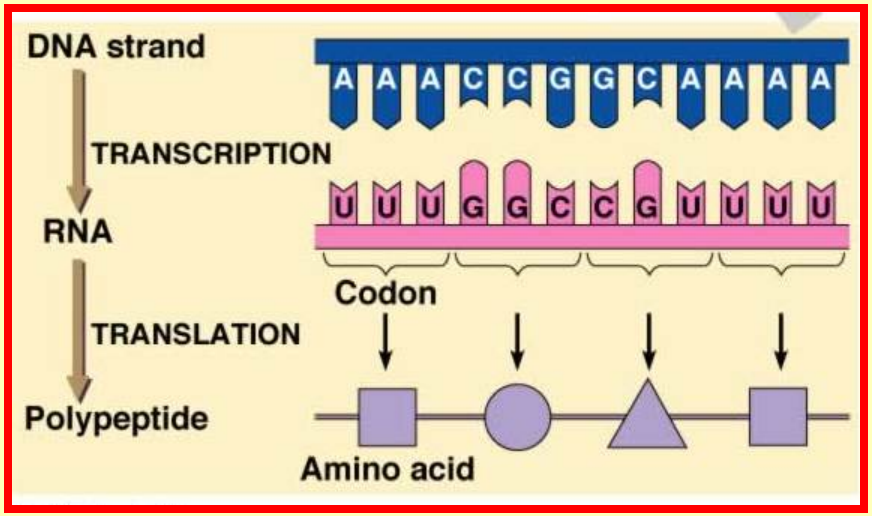
b. **Small cytoplasmic RNA (scRNA)** which participates in post-translation protein trafficking and targeting in the cell.

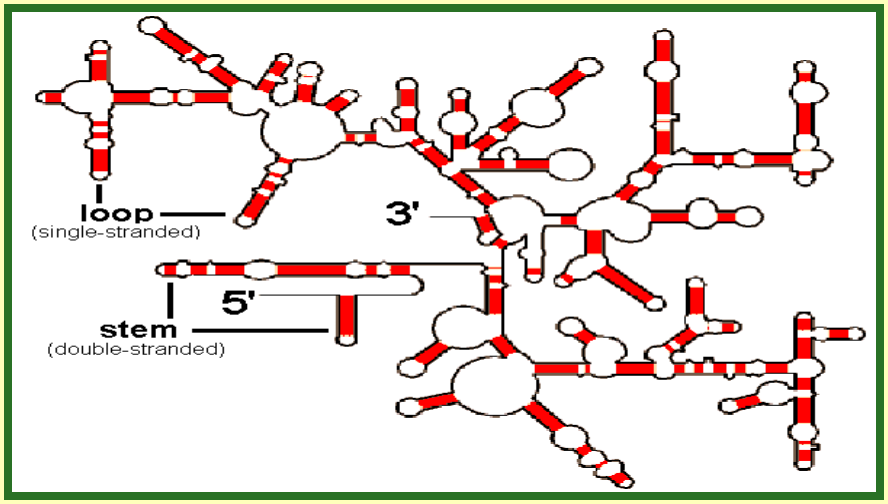
c. **Ribozymes** or catalytic RNA molecules with specific enzymatic activities.

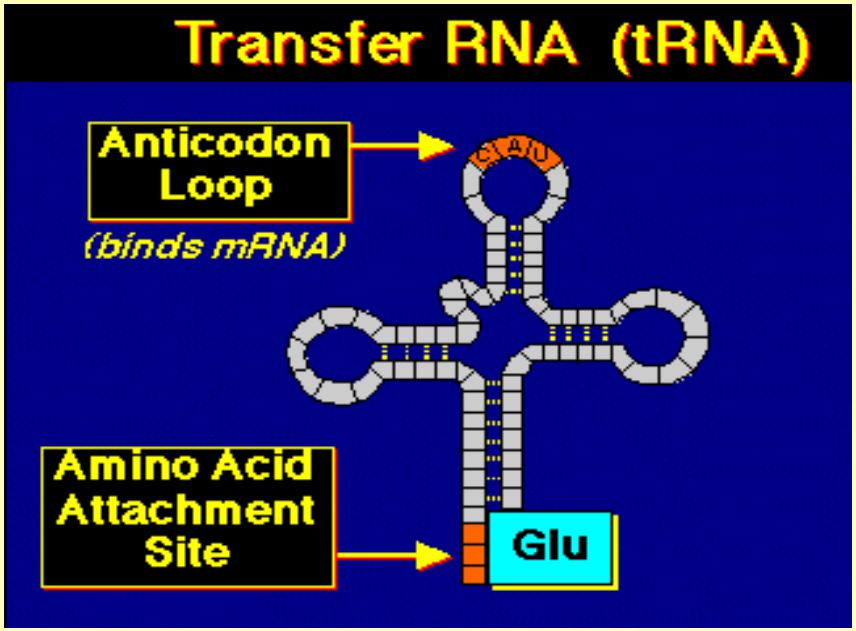
MicroRNA or small RNA (miRNA) are post-transcriptional regulators that bind to target messenger RNA transcripts (mRNAs), usually resulting in gene silencing and decreased transcription. miRNAs are short ribonucleic acid molecules, consisting, on average, of 22-24 nucleotides long. The human genome may encode over 1000 miRNAs which may target about 60% of mammalian genes and are abundant in many human cell types. Each miRNA may repress hundreds of mRNAs. miRNAs are well conserved in eukaryotic organisms and are thought to be a vital and evolutionarily ancient component of genetic regulation.

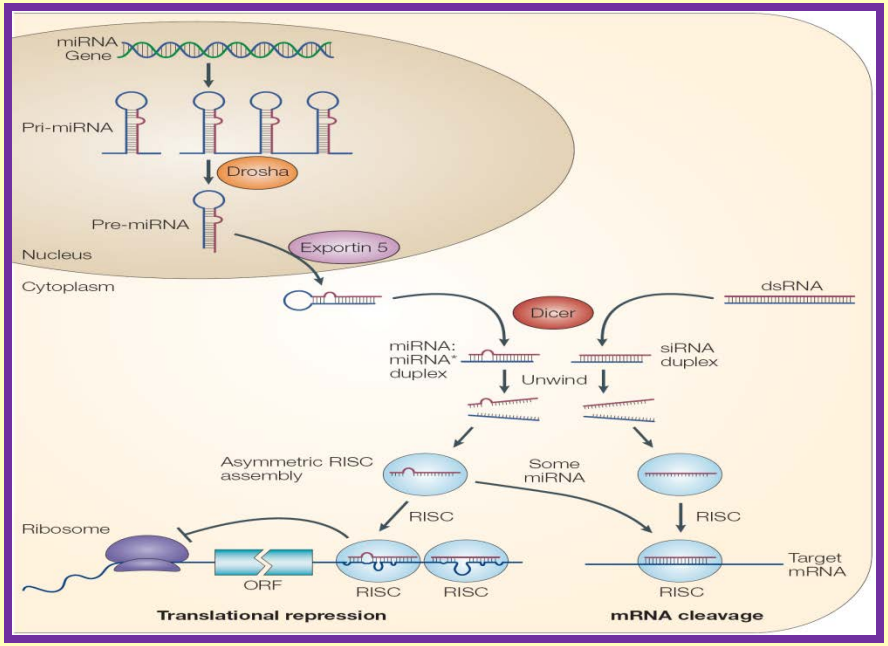
Some types of miRNA (Guide RNA) have a critical role in correcting some defects in mRNA (mRNA repair) or inducing structural defects in mRNA leading to defective translation and disease in spite of normal gene structure (secondary mutation). This effect partly explains the finding of normal structure of genes in some disease conditions known to result from mutations of these genes.

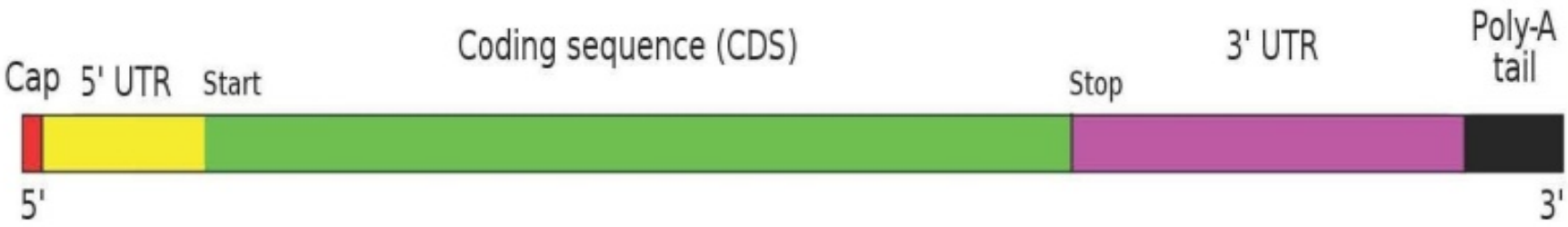
Functional Categories Of RNAs











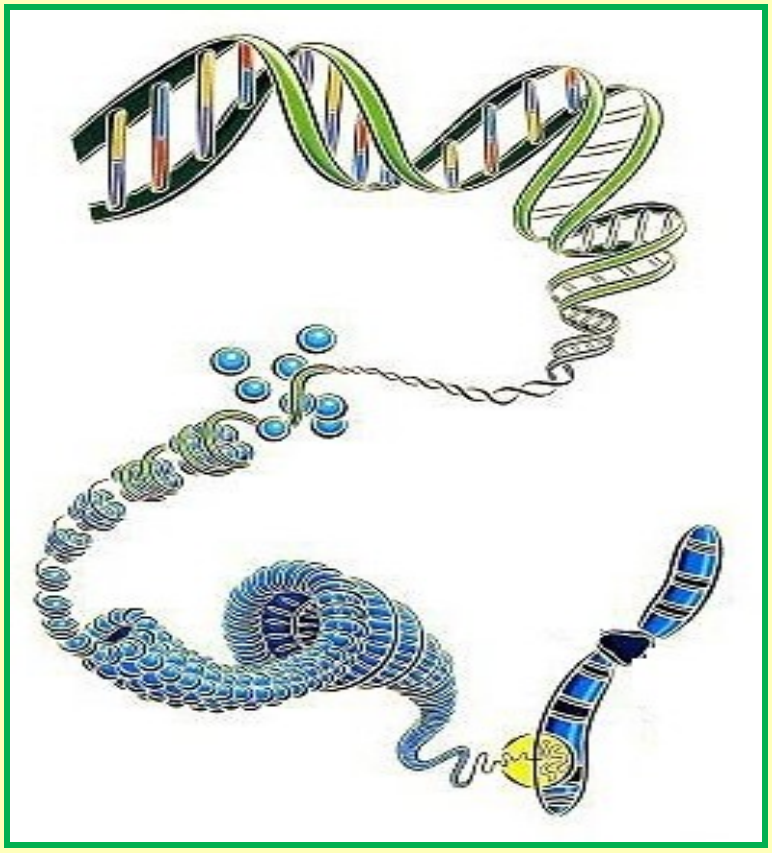
A fully processed, final mRNA consists of a 5’ cap of 7 methyl-guanosine at its 5' end, 5’ UTR (untranslated region), coding region consisting of spliced exons, 3’ UTR (untranslated region) and a poly (A) (polyadenylate) tail at its 3' end

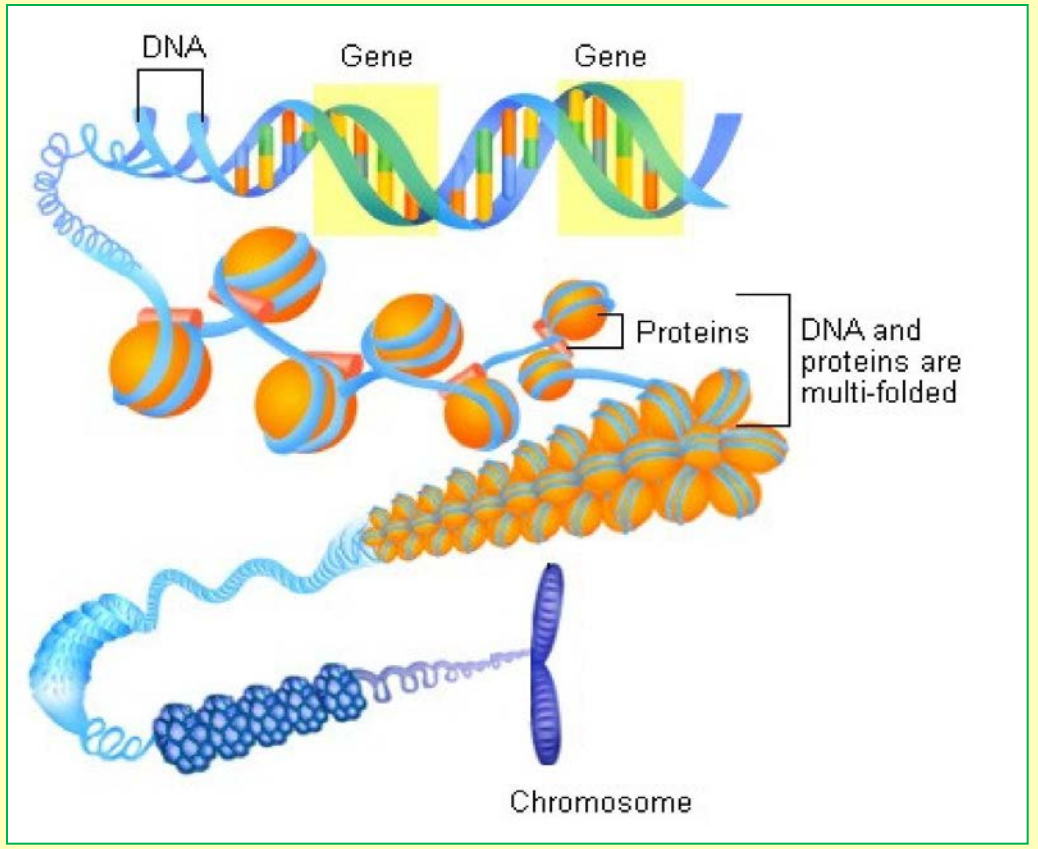
Organization OF The Genetic Material

The genome, or the sum total of the genetic material in the cell, consists of genes in addition to other non-genic or gene-related components. Each species has its own specific genome that differs from the genome of any other species as regards the number of genes, their cellular distribution and the size of the genome itself, among many other inter-species differences.

In human cells, the human genome is unequally distributed into a major part, constituting more than 99.999 % of its size, organized in the form of long strands, open-ended chromosomes contained in the nucleus and referred to as the nuclear genome which comprises between 25000 – 38000 genes distributed over the chromosomes.

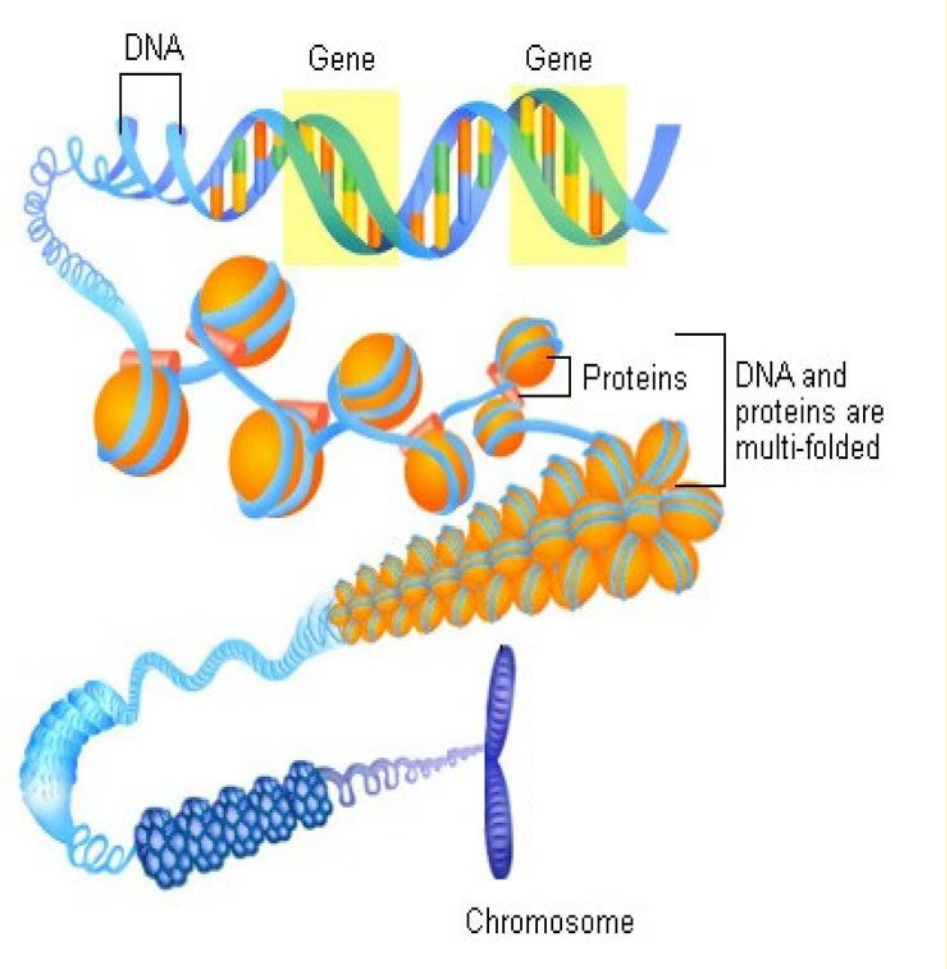
Each chromosome consists of a very long double stranded molecule of DNA wrapped with a heavy coat of basic proteins composed mainly of histones and protamines. These DNA-associated proteins offer support and protection to the DNA and play a critical role in regulating many aspects of gene functions.



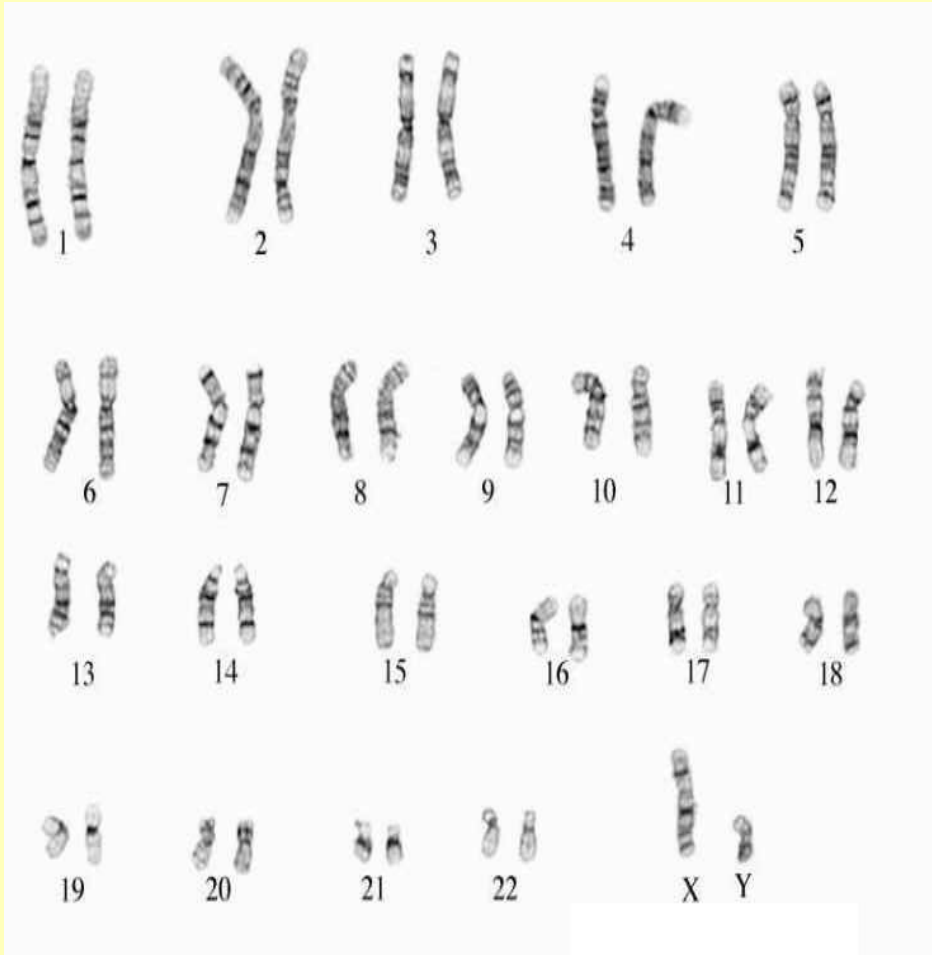


Structure & Types Of Chromosomes

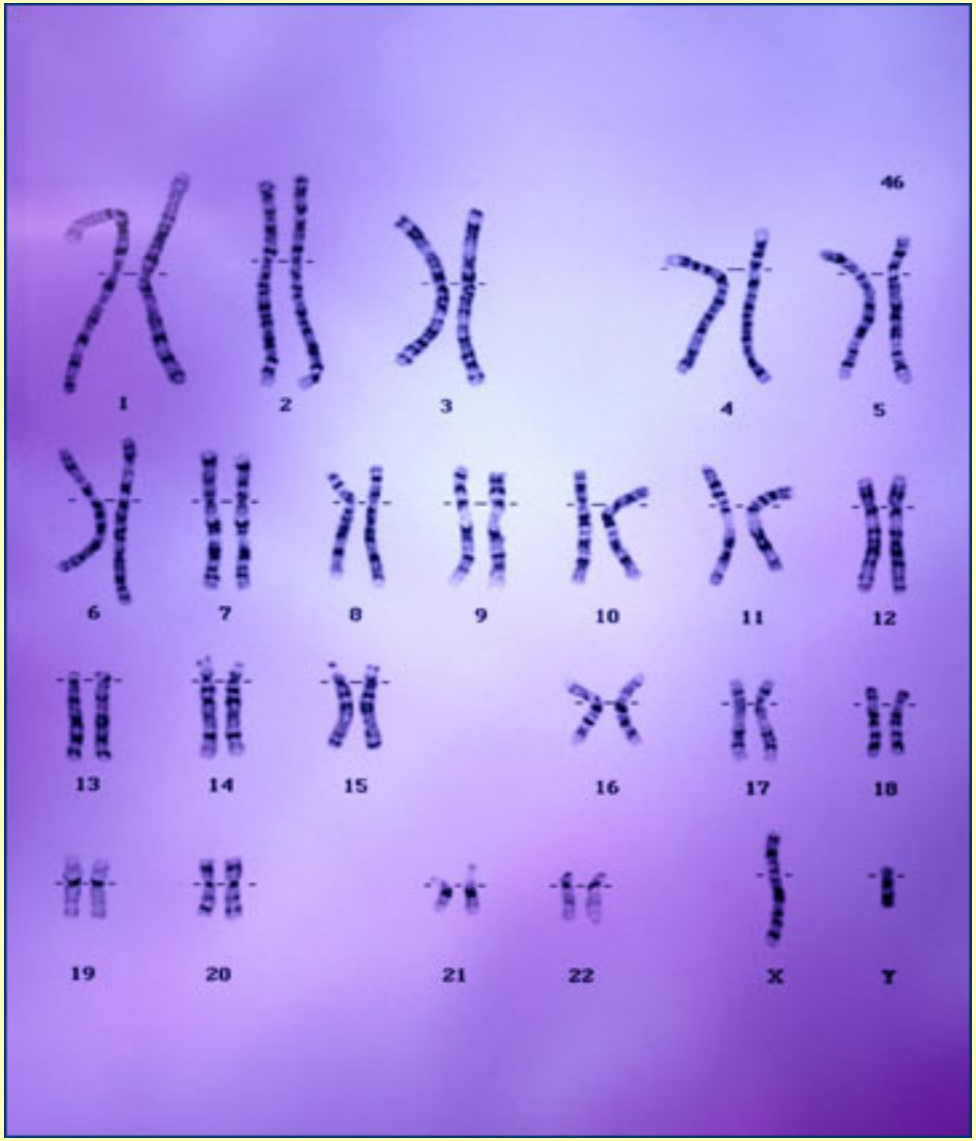
Structure Of Chromosomes



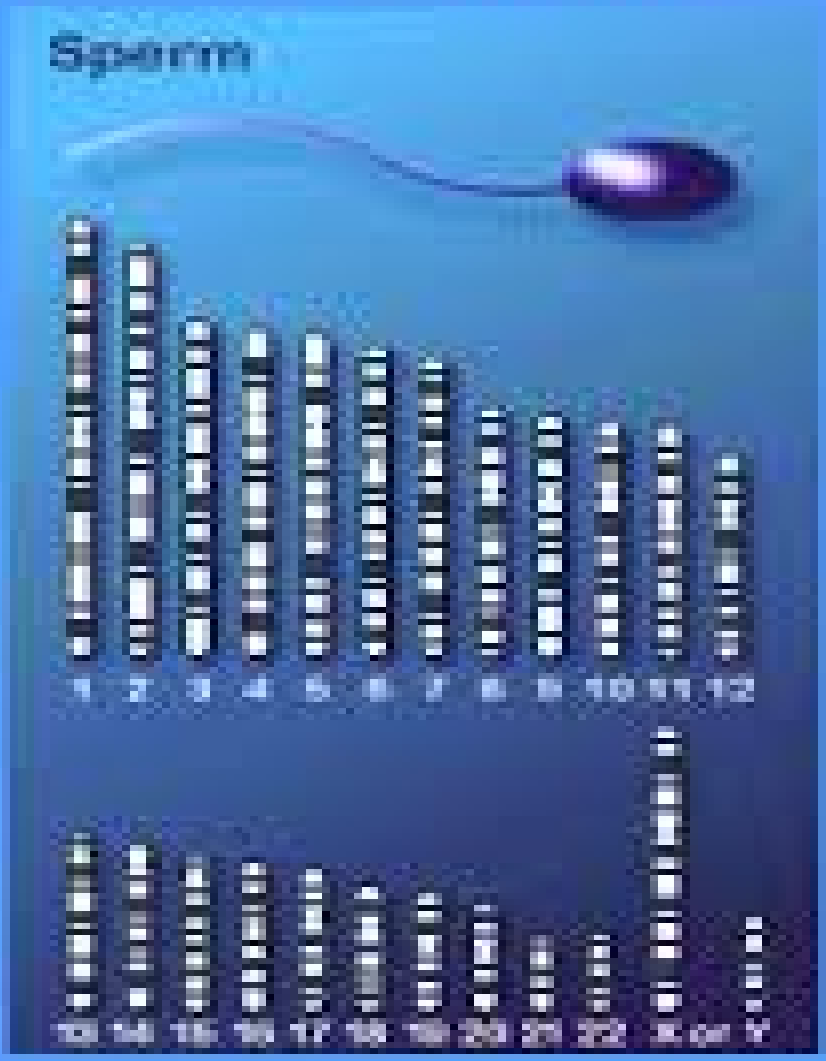
Types Of Chromosomes



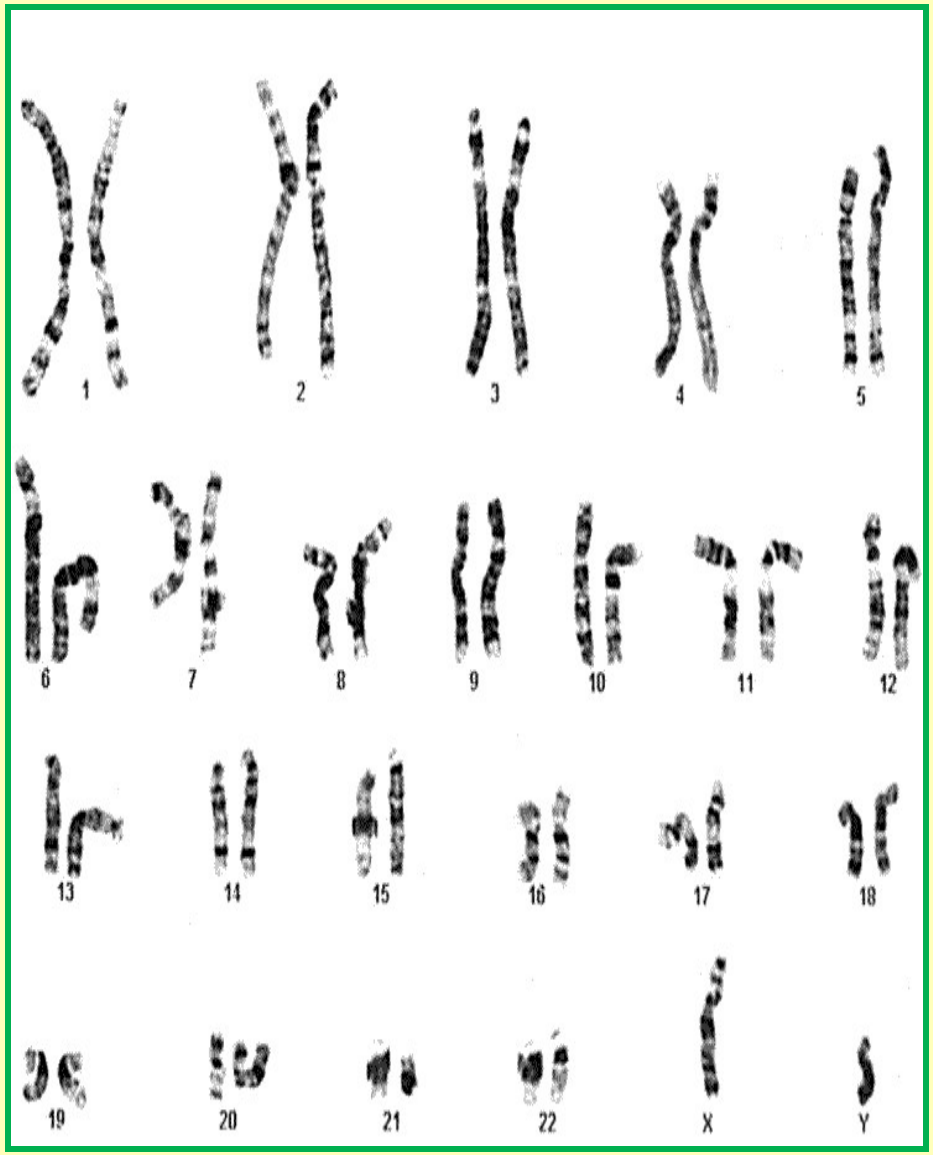
Organization Of The Nuclear Genome As Chromosomes

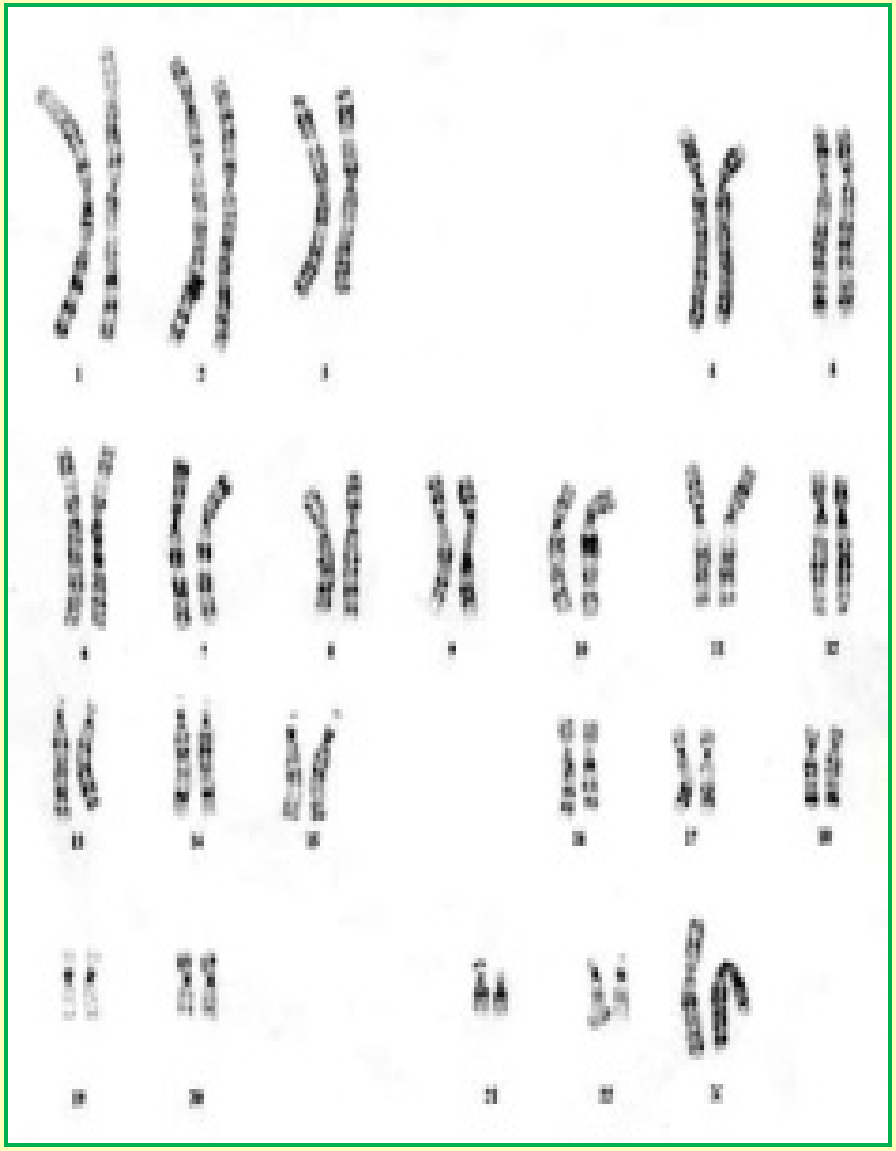


Haploid Genome

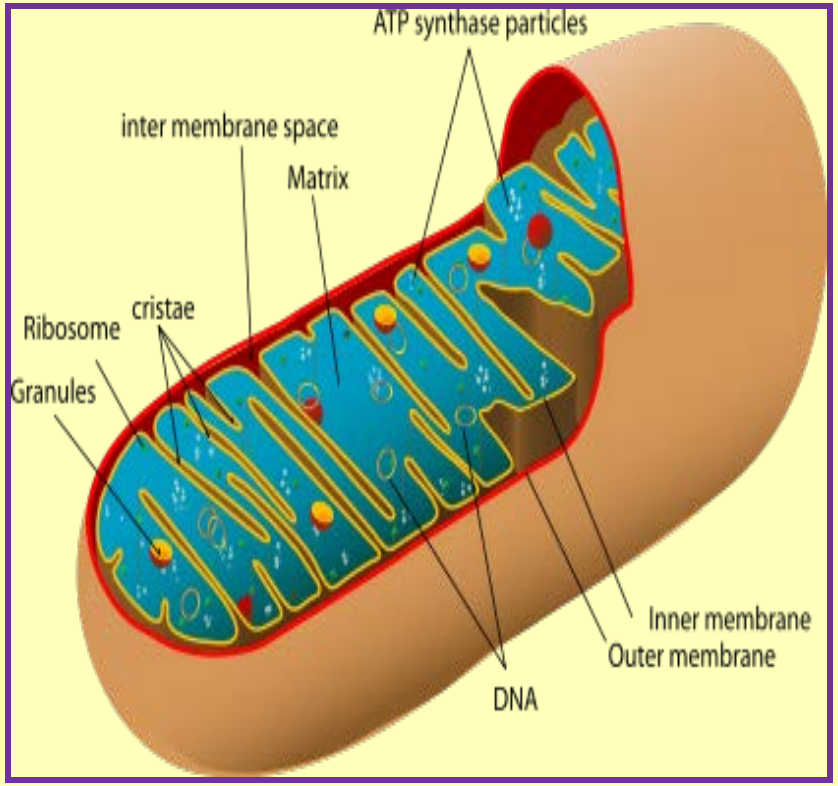


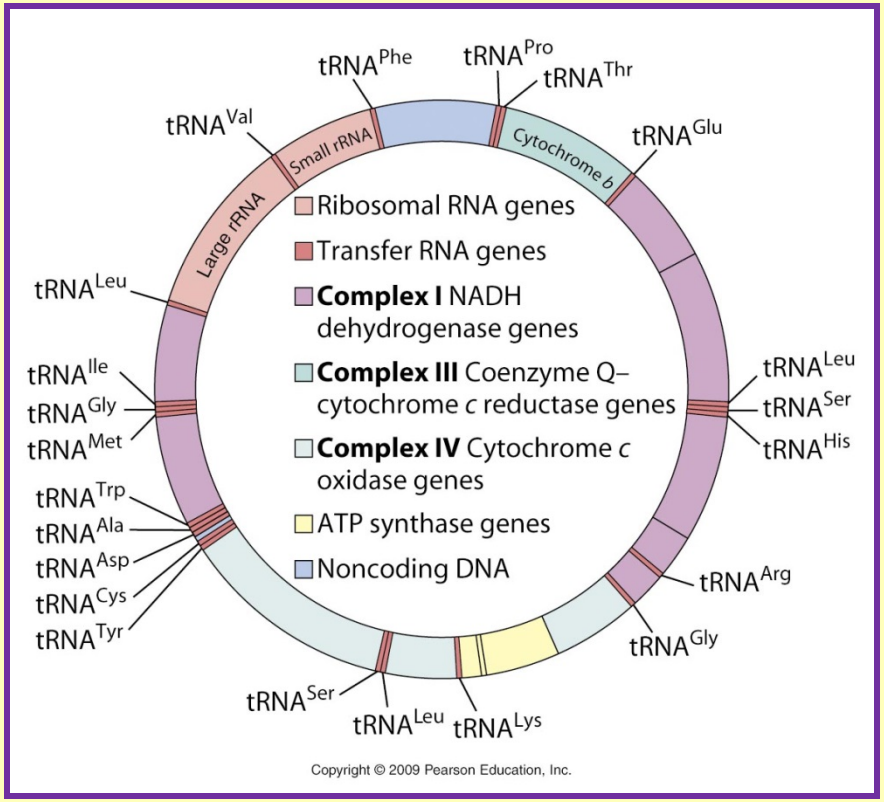
Normal Male (46,XY) And Female Karyotypes (46,XX)



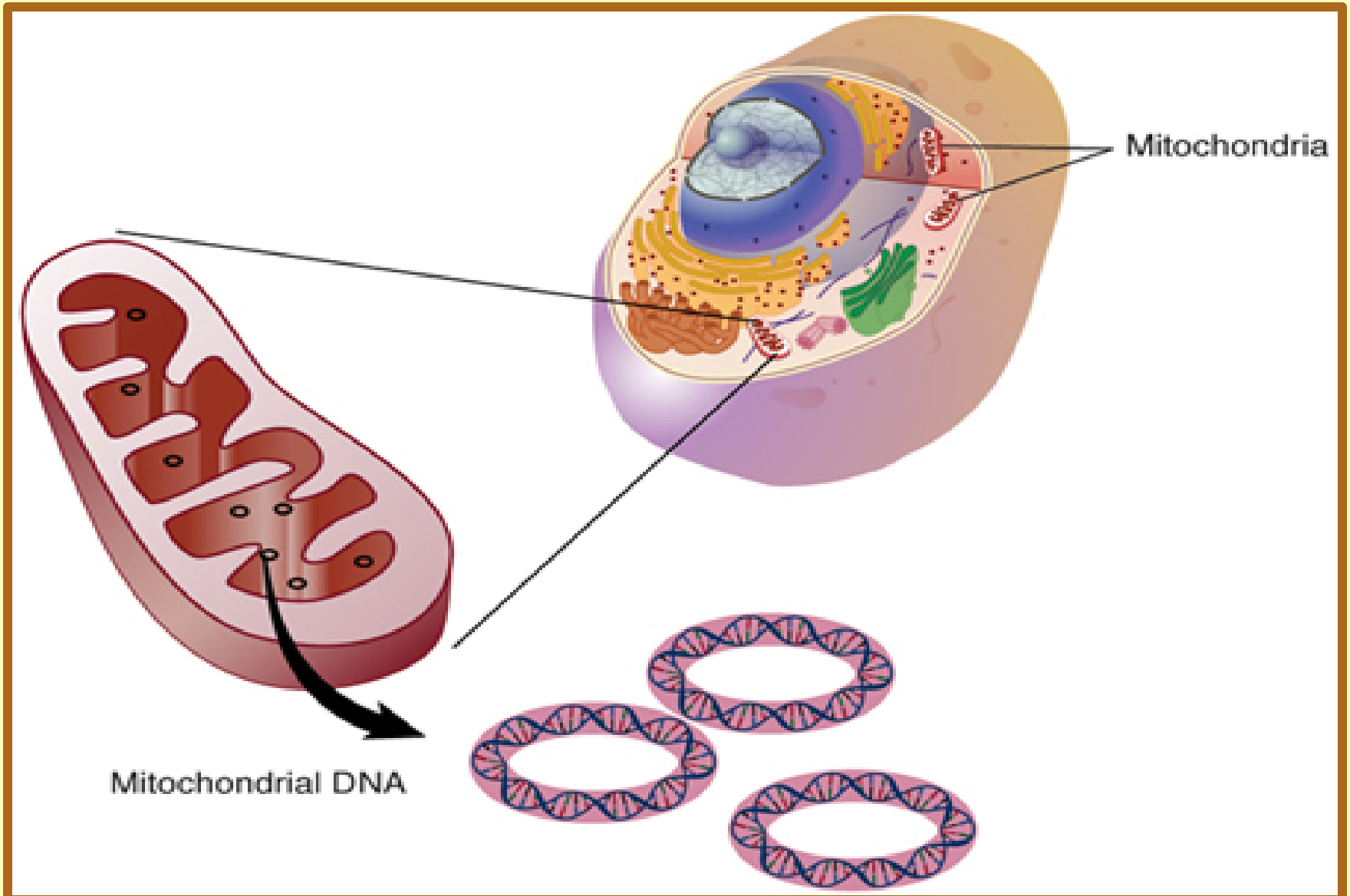


The remaining tiny part of the human genome exists in the form of varying numbers, tens to thousands, of very small closed circular double stranded structures present inside the mitochondria and is referred to as the mitochondrial genome. Each molecule of the mitochondrial genome (mtDNA) consists exclusively of 37 genes





Mitochondrial DNA



Though it constitutes a very tiny fraction of the whole genome, mtDNA is indispensable for life because it codes for proteins that mediate ATP production in the cell in addition to many other important functions like apoptosis and many other vital metabolic activities like lipid oxidation and steroid biosynthesis.

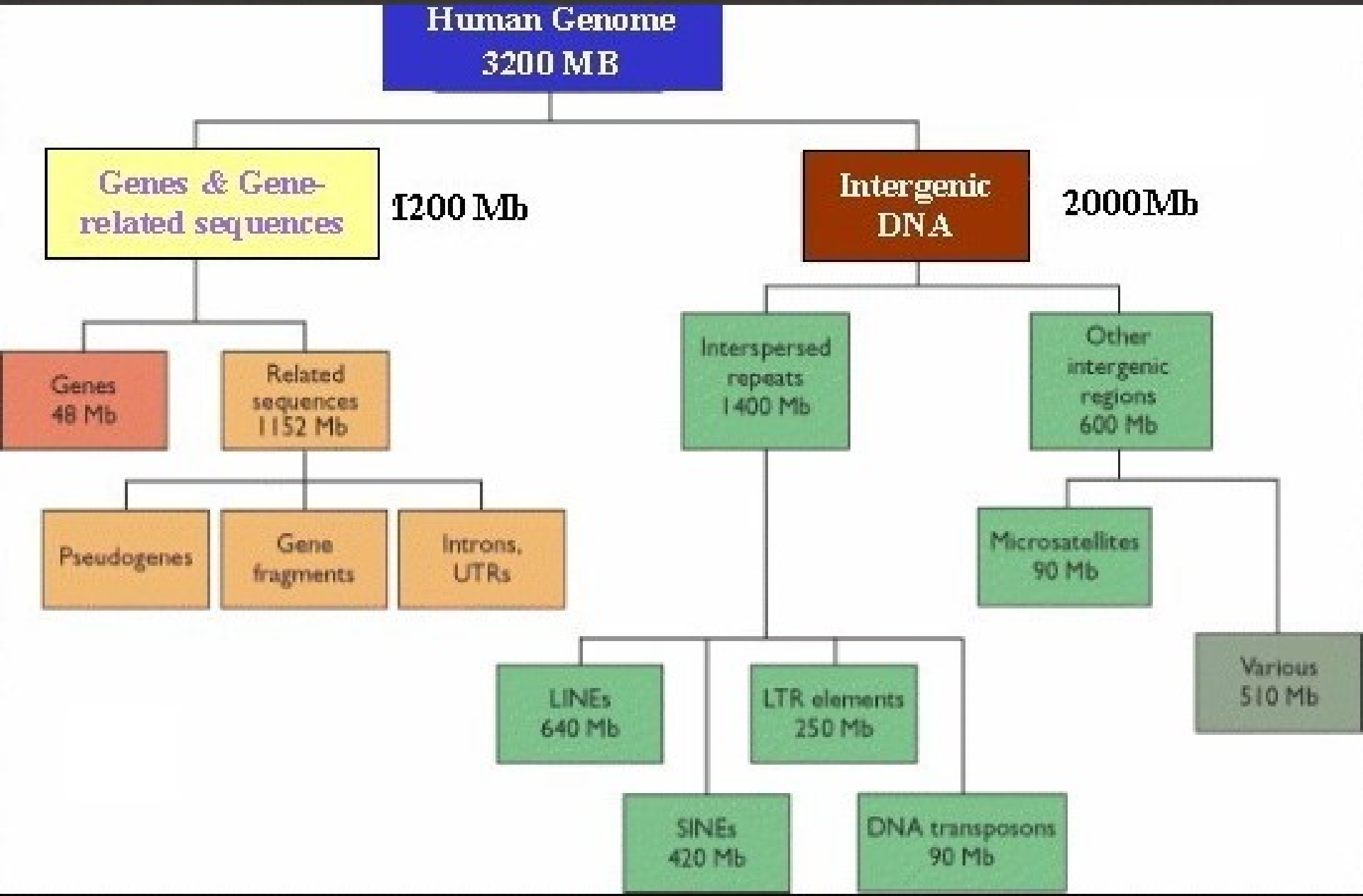
The number of mitochondria and the number of mtDNA molecules in each mitochondrion varies according to the metabolic activities of the cell. The most active and energy-demanding cells, like neurons, heart muscles, the retina, skeletal muscles, endocrine glands, kidney cells and liver cells have the largest numbers of mitochondria within their cytoplasm and the largest numbers of mtDNA molecules in each mitochondrion as well.

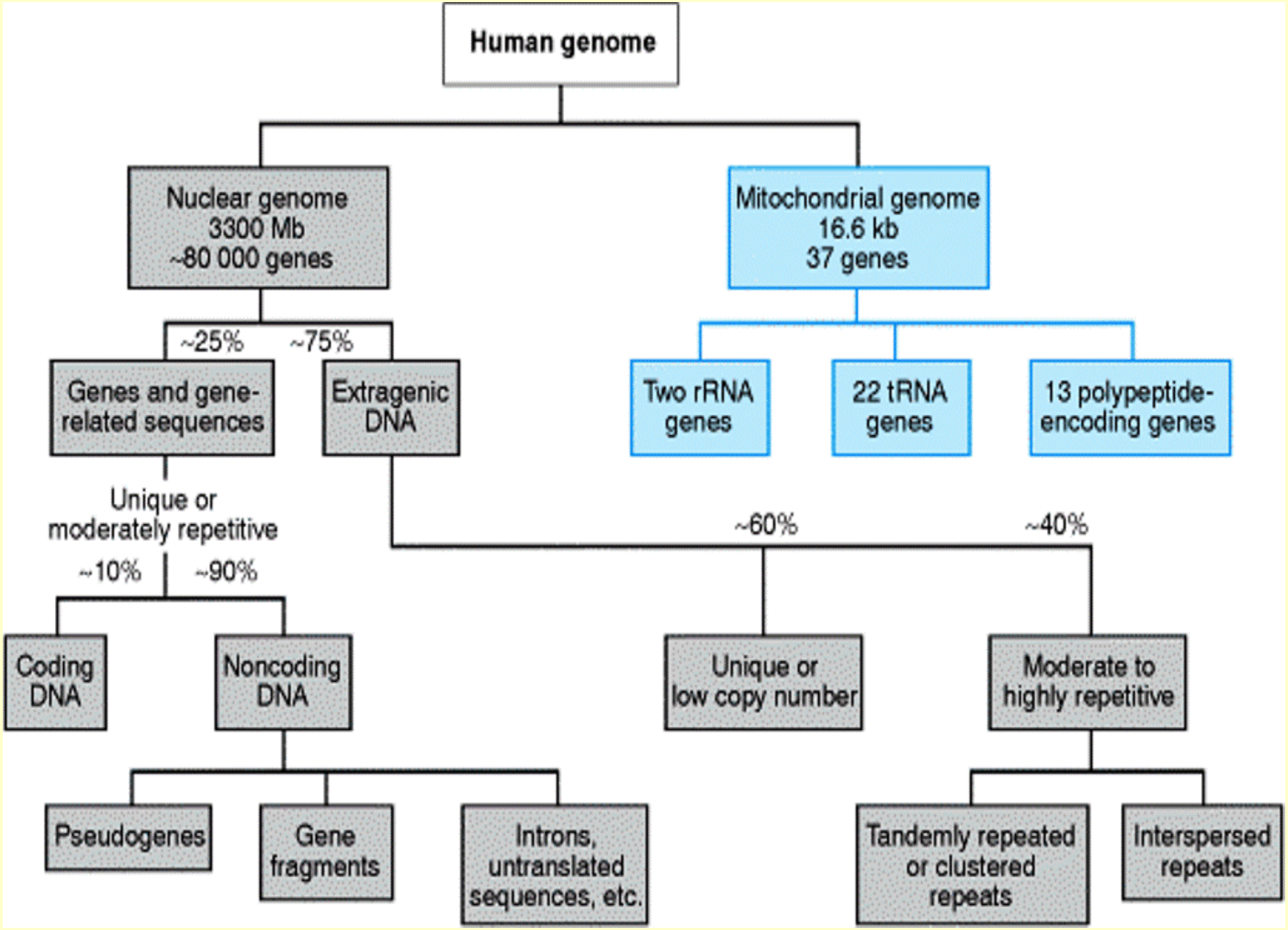
The nuclear genome in each human germ cell, ovum and sperm, is organized into a set of 23 separate chromosomes known as the haploid genome which represents the unit genome of humans.

Upon fertilization, both haploid genomes of the sperm and the ovum constitute a diploid genome consisting of their 46 chromosomes that characterizes the nuclear genome of the zygote as well as of all somatic cells descendant from it.

With very few exceptions, the sperm does not contain mitochondria. Nearly all mitochondria, and hence the mitochondrial genome, present in the zygote and in all body cells are descendent from the mitochondria present in the ovum at fertilization.

Organization of the human genome




Human genome
Nuclear genome
3300 Mb
~80 000 genes
Mitochondrial genome
16.6 kb
37 genes
~25%
~75%
Genes and gene-related sequences
Extragenic DNA
Two rRNA genes
22 tRNA genes
13 polypeptide-encoding genes
Unique or moderately repetitive
~10%
~90%
~60%
~40%
Coding DNA
Noncoding DNA
Unique or low copy number
Moderate to highly repetitive
Pseudogenes
Gene fragments
Introns, untranslated sequences, etc.
Tandemly repeated or clustered repeats
Interspersed repeats

Structural/Functional Organization of the Human Genome

Spatial Organization of the human Genome

A. Nuclear Genome (Chromosomal/nDNA)

B. Mitochondrial Genome (mitDNA)

C. Cell membrane-associated DNA (cm-DNA)

A tiny portion of DNA in the cytoplasm of somatic cells attached to the inner side of the cytoplasmic membrane. It has physical and chemical properties different from both chromosomal and mitochondrial DNAs and represents a portion of the heterochromatin of the centromeric and peri-centromeric regions of chromosomes that exited to the cytoplasm.

cm-DNA is transcribed separately in the cytoplasm by a specific RNA polymerase different from that used for nuclear DNA transcription.
The potential functions of cmDNA are largely undefined. However, many putative roles have been assigned to it including mediation of cellular activities, e.g. control of signal transduction in the cytoplasm and induction or regulation of apoptosis. Inappropriate over-transcription of cmDNA might result in disturbance of the intricate balance between oncogenes and tumor suppressor genes and has been implicated in development of some malignancies, like breast cancer.

Functional Organization of the Human Genome

A. Master Genes

Function continuously, examples include:

Energy (ATP) production genes

Cell cycle/cell division regulator genes

DNA repair genes/Chromatin assembly genes

Cytoskeleton stabilizing genes

Cellular transport regulator genes.

B. Regulatory Genes

Sensors of gene stimulator/gene suppressor signals to switch on / switch off coding genes through synthesis of transcription regulatory factors.

C. Structural, Protein Coding Genes

D. Structural, RNA Coding Genes

E. Non-Coding Regions

Transposons

Transposons represent a unique feature of the genome of most living creatures. They represent one type of mobile genetic elements (MGE) which are sequences of the genome that can move from their original locations to other sites within the genome or make a copy of their sequence to be inserted in other parts of the genome. Other mobile genetic elements include plasmids, group II catalytic introns or ribozymes and bacteriophages. The movement of transposons from their site to another site happens via one of two mechanisms: the replicative mechanism and the conservative mechanisms. In the replicative mechanism, the tranposon element replicates making a copy of itself and the new copy gets inserted in a new site of the genome thus leading to duplication of the transposon sequence.

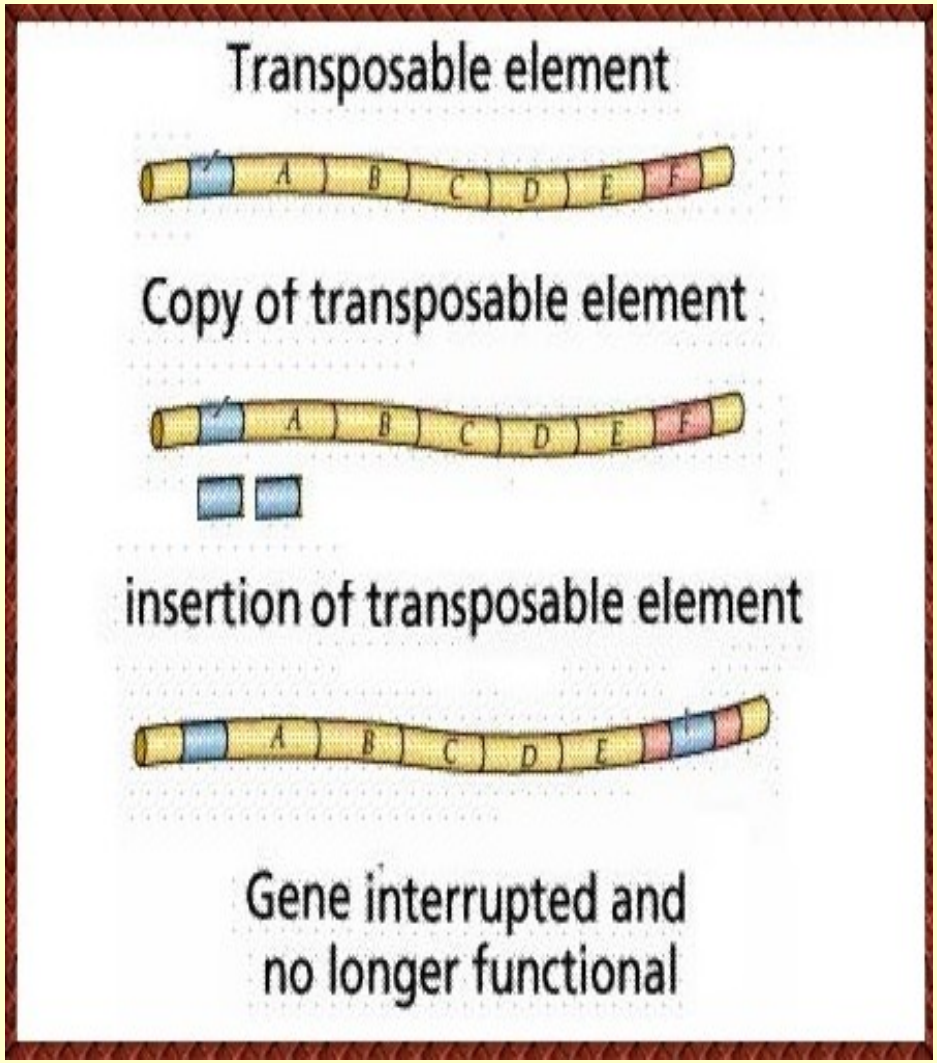
This category of transposons is named Class I Retrotransposons. The second group of transposons, Class II DNA Transposons, follows the conservative mechanism where the transposon detaches and moves, or transposes itself, from its original location on a specific chromosome to a new site on the same or on a different chromosome. In either case, insertion of a new segment into a normal segment leads to disruption of the integrity of the normal sequence. This phenomenon, referred to as insertional mutagenesis, is shared with many retroviruses. The resultant effect of this process depends on the site of insertion of the mobile element. If it occurs within nonfunctional inter-genic portions of the genome no harm is to be expected.

However, insertional mutagenesis within functional genes results in variable degrees of damage depending on many factors, like the size and the site of the new insertion. It usually results in interruption of structural integrity of the gene and constitutes one major factor that underlies the occurrence of spontaneous mutations of the genetic material and pathogenesis of genetic disorders. On the other hand, creation of new genetic combinations between receptor segments of the genome and inserted transposons might be considered, within the context of evolutionary genetics, as one mechanism for genomic diversity and phenotypic evolution if these new genetic combinations result in construction of new functional genetic elements and begin their own expression.

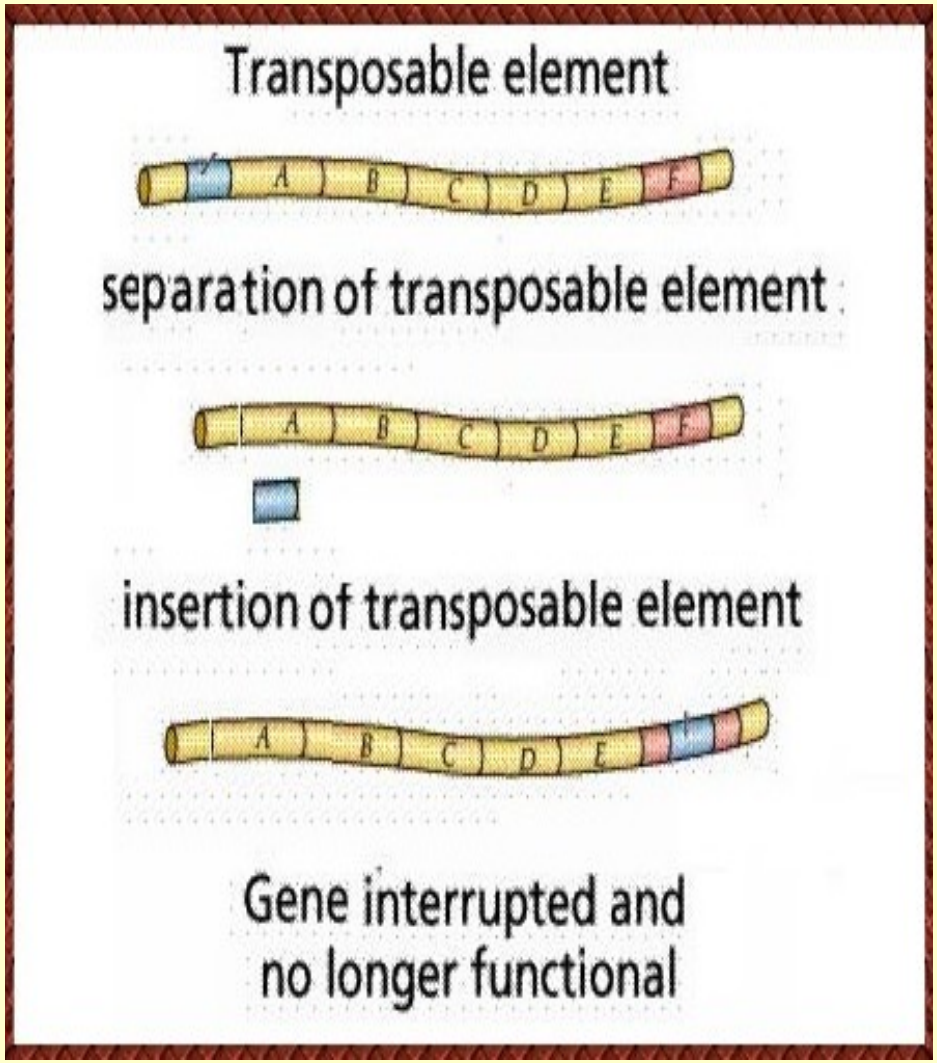
Formation of new metabolic networks and the de novo creation and establishment of novel organized regulatory pathways are two possible assumed mechanisms that can underlie the acquisition of new genetic phenotypes due to transposon activity. In bacteria and some other lower eukaryotes, some transposable elements contain genes coding for proteins products, mostly enzymes, needed by the bacteria, e.g., antibiotic resistance genes) to combat effects of antibiotics.

TYPES OF TRANSPOSONS

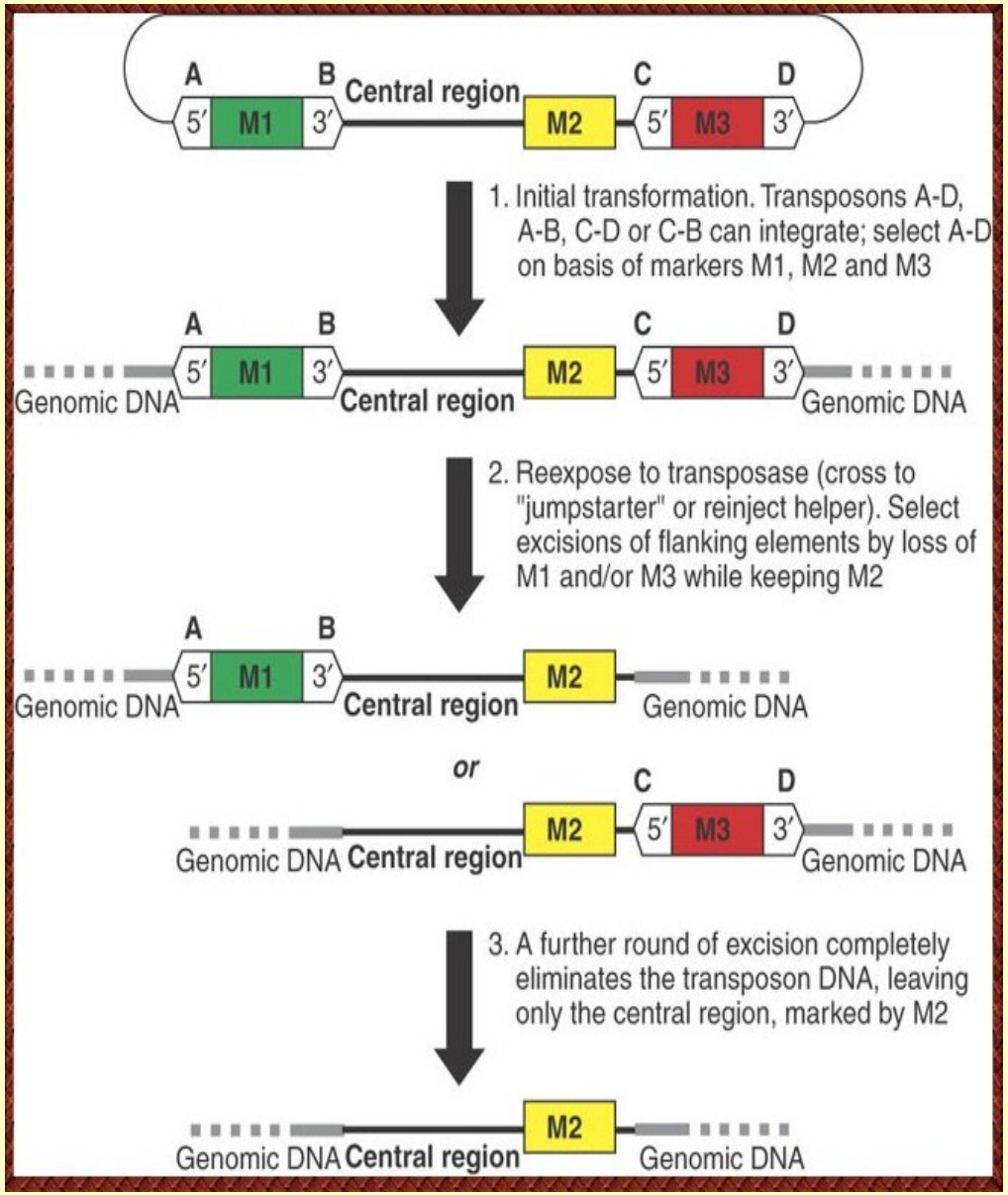
Class I Retrotransposons

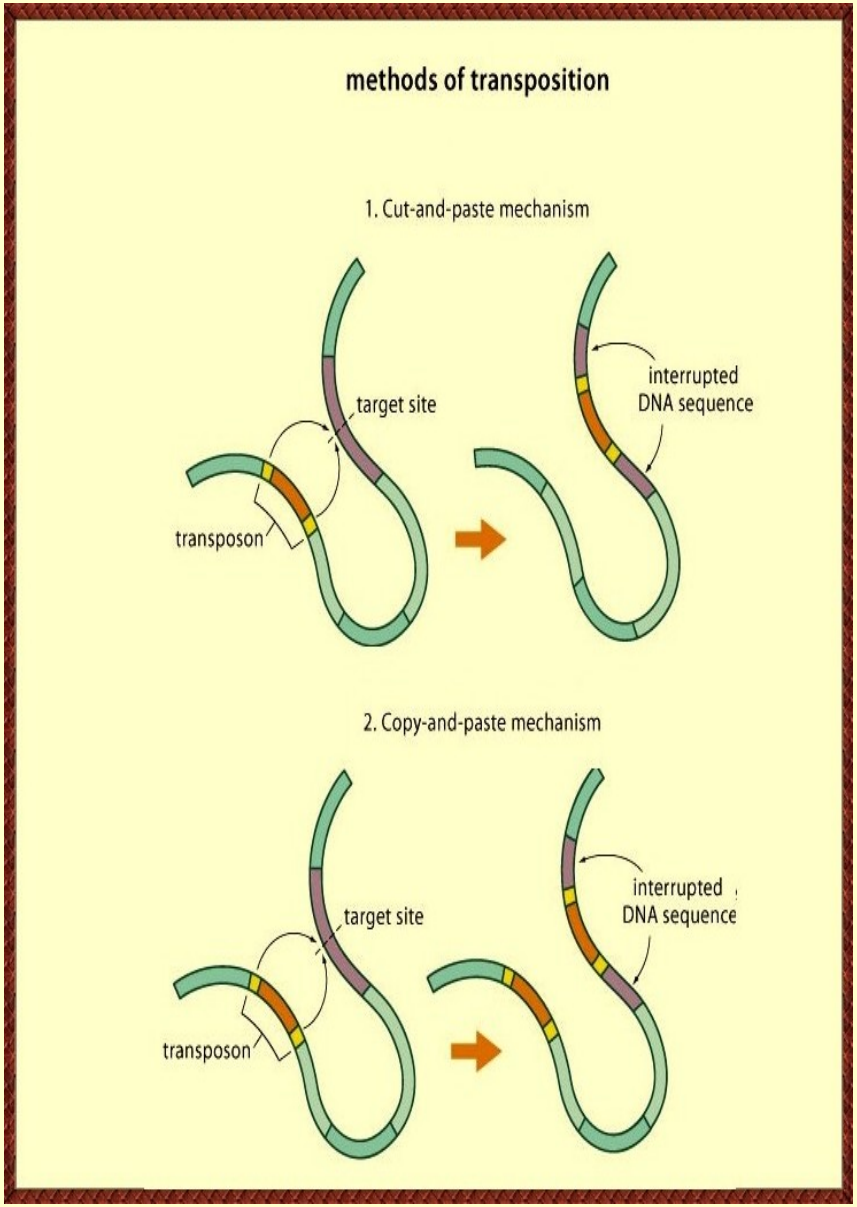


Class II DNA Transposons



Methods & Mechanisms Of Transposition





Pseudogenes

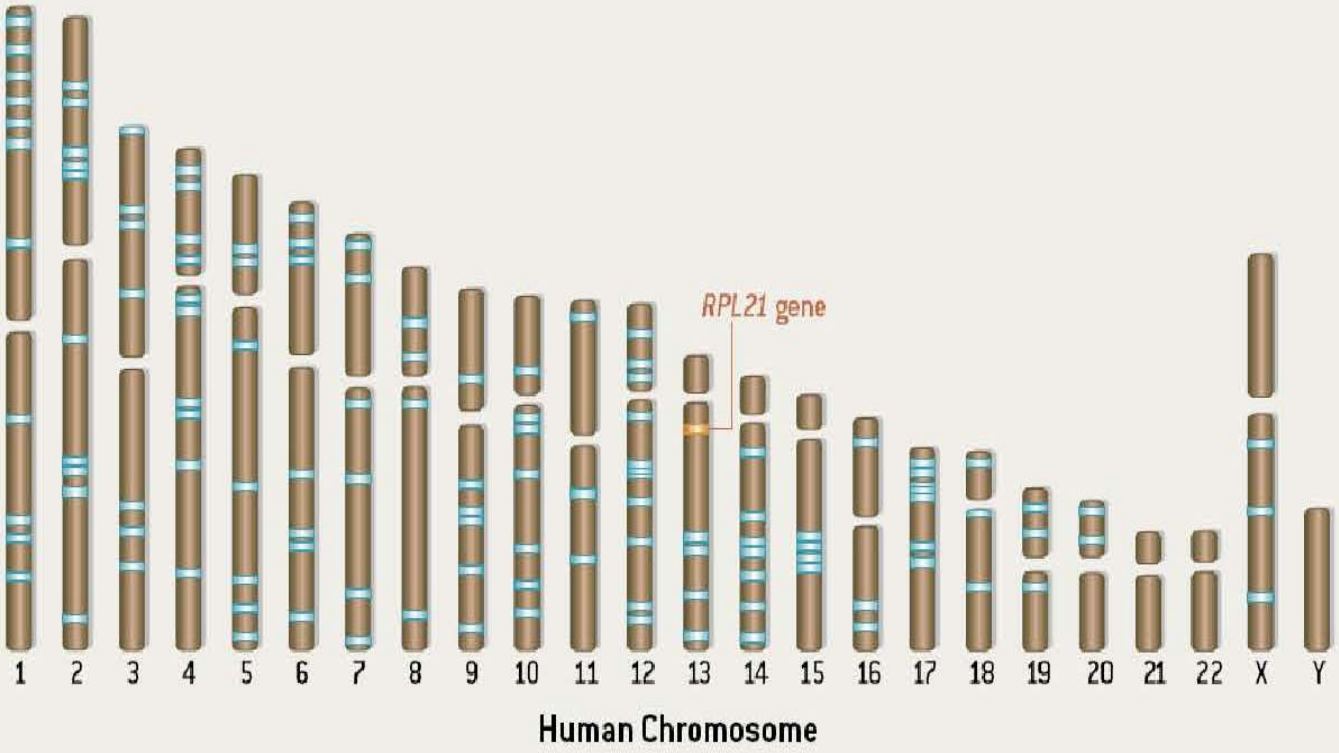
Pseudogenes are DNA sequences that structurally resemble functional genes. There are two types of pseudogenes known as processed and unprocessed pseudogenes. Processed genes are found on different chromosomes, they lack introns and certain regulatory elements, often terminate in adenine series, and are flanked by direct repeats. They may be complete or incomplete copies of genes or mixtures of several genes. They are believed to have occurred through transcription of the original gene to mRNA followed by posttranscriptional removal of introns and forming back DNA through a reverse transcription process. Unprocessed pseudogenes, having their original introns and associated regulatory elements, usually exist as clusters of similar structural sequences on the same chromosome.

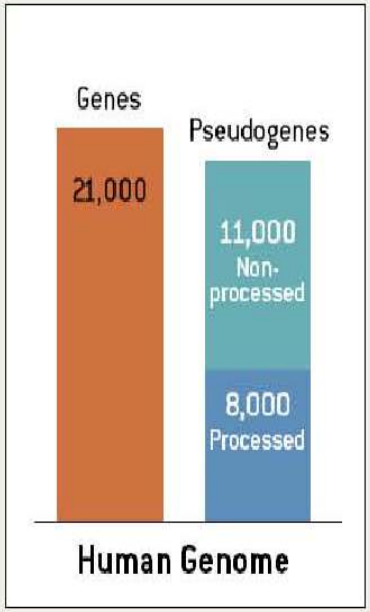
Their active expression is usually suppressed by one or more type of point or small mutation affecting its promoter area, including deletion, insertion or change to stop or termination mutation. Unprocessed pseudogenes are believed to have arisen by duplication of their parent gene. Pseudogenes may have similar role in cases of catastrophic unrepairable damage of the original genes. They might function as standby genes ready for repair and/or reactivation to undertake the functions of damaged genes. Also, they may have quantitative enhancing roles during early embryonic growth and differentiation where the needs for gene products are particularly most demanded by fast growing and dividing cells during this stage of development.

In many instances, pseudogenes code foe proteins/small regulatory-interfering RNA that regulate functions of tumor suppressor genes and oncogenes.

Many Pseudogenes affect the functions of functional genes by decreasing mRNA stability and expression of these genes.

Pseudogenes in Nuclear Human Genome





PSEUDOGENE DESCENDANTS (*blue*) of the ribosomal protein gene *RPL21* (*orange*) are scattered across the human chromosomal landscape. Overall distribution of pseudogenes in the human genome appears to be completely random, although some local genome regions tend to contain more pseudogenes. Those DNA regions may be analogous to certain geochemical environments that better preserve mineral fossils. Identification of genes and pseudogenes is an ongoing process, but to date more than 19,000 pseudogenes have been identified in the human genome—only slightly less than the current tally of around 21,000 human genes (*inset*). About 8,000 of our pseudogenes are processed; the rest include duplicated pseudogenes and other nonprocessed subcategories.

Pyknons

Pyknons are short non-coding DNA sequences about 20-22 nucleotides in length. They are widely distributed in the nuclear human genome in both the inter-genic and intronic regions of the genome, constituting about 1/6th of the human genome. This makes them the most frequent, variable-length DNA sequence motifs in the human genomes. Pyknons have a remarkable degree of structural conservation. Their presence in the 3' UTRs (un-translated regions) of genes may indicate a potential regulatory role in posttranscriptional processing and modifications of mRNA. Though they do not share in either protein synthesis or RNA transcription, pyknons are functional genetic elements associated with mediation of specific biologic cellular processes. They are putative factors implicated in susceptibility to some common human genetic disorders.

Disturbed genomic regulation of function(s) of pyknons might underlie the development of this genetic susceptibility. The considerable size of pyknons in the genome coupled with the intimate functional relationship between them and many subtypes of microRNAs suggest a pivotal role played by pyknons, probably as global regulators of gene function. Some unique sequences of human genomes are designed from series of short template octamer sequences which are embedded into pyknon's sequences and represented by hundreds (up to thousands) of copies in a genome.

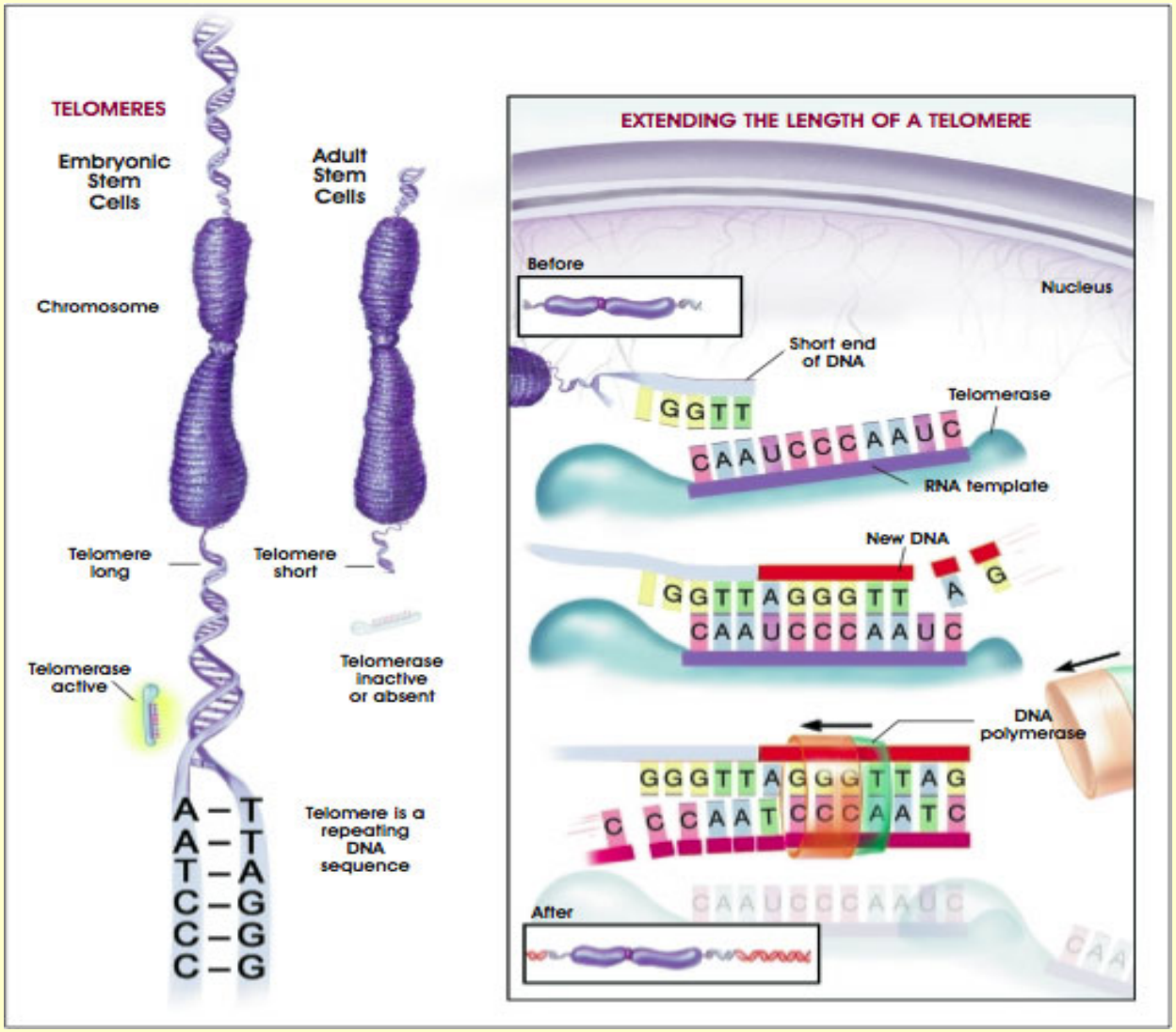
The assumed regulatory roles of pyknons might be exerted via different mechanisms. The small nucleotide number of pyknons, similar to the small nucleotide number of most microRNAs species, elicits some questions regarding their origin in the genome, and the possibility that pyknons might represent non-classic genes or transcriptional units capable of directing and regulating synthesis of some microRNAs species for specific biological activities.

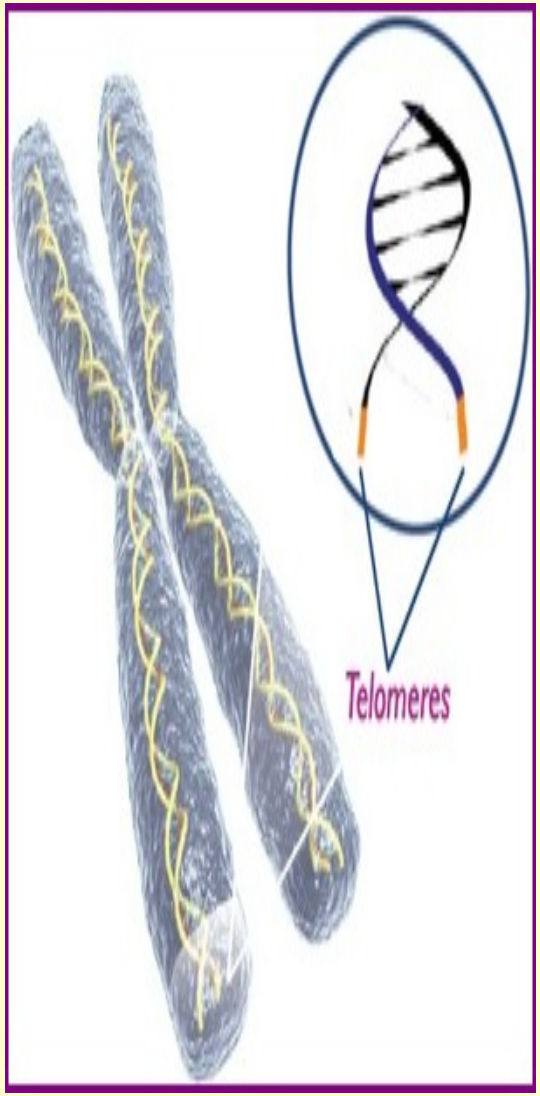
Pyknons are short blocks from the noncoding parts of the human genome present within nearly all known genes and relate to many important biological processes.

Telomeres (terminal/interstitial telomeric sequences)

Telomeres are specific noncoding, repetitive nucleotide sequences consisting of as many as 2000 repeats of the sequence (5' TTAGGG 3') located at the ends of linear chromosomes of most eukaryotic organisms. They protect the chromosome ends from being fused to each other and from fraying upon exposure to damaging agents. Over time, however, each cell division cycle results in loss of a portion of the telomere sequence leading to progressive shortening of the telomere because DNA replication cannot continue their duplication all the way to the end of chromosomes.

Telomeres in Nuclear Human Genome





If cells divide without telomeres, they would lose the ends of their chromosomes, and the important genetic information they contain. In human blood cells, the length of telomeres ranges from 8,000 base pairs at birth to 3,000 base pairs as people age and as low as 1,500 in elderly people. Each time a cell divides, an average person loses 30 to 200 base pairs from the ends of that cell's telomeres.

Consumption of telomere portions during cell division is partly corrected by re-synthesis by a specific enzyme named telomerase reverse transcriptase. This enzyme is found only in certain types of cells which comprise germ line cells, embryonic stem cells and adult stem cells including cancer stem cells and their progenitor cells.

The activity of the enzyme in these cell types explains many aspects of their biologic potentials. Prolonged and persistent synthesis of telomerase reverse transcriptase enzyme is a constant feature of most cancer cells and represents an important malignant phenotype of these cells underlying their ability to grow and divide indefinitely.

However, at a certain stage of somatic cell life cycle, no more telomere sequences could be lost and gradual deterioration of chromosome integrity ensues, leading ultimately to replicative senescence, enhanced aging and cell death. The role played by telomeres in regulating the number of cell divisions during the life span of the cell, referred to as the Hayflick limit, reveals the critical role played by telomeres in keeping genomic integrity and stability within safe functional limits all through the life span of the cell.

Though telomeres are found exclusively at the ends of linear chromosomes, interstitial telomeric sequences (ITSs) with their specific repeats of (5' TTAGGG 3') are found scattered throughout the human genome,

particularly within the middle of chromosome 2 which contain pre-telomeric sequence, a telomeric sequence, an inverted telomeric sequence and an inverted pre-telomeric sequence. ITSs are often functionally important to the genome. The chromatin organization of telomeres can silence genes and has been linked to epigenetic modes of inheritance. Furthermore, difference classes of transcripts are derived from telomeres and their flanking repetitive DNA regions. These are involved in numerous cellular and developmental functions. It seems more likely that ITSs are sites where TTAGGG repeats have simply been added to chromosomes by telomerase enzyme and that many of these ITS sites are associated with distinct sets of proteins which have been linked to important functional roles within the genome, such as recombination hotspots.

Accumulating observations indicate that telomeres have important potential roles in many critical cellular processes. These processes include control of cell division, regulation of cell longevity, apoptosis, maintenance of optimal performance of stem cells and progenitor cells during early development and ensuring proper genomic replication of germ line cells during gametogenesis, among many others. The finding of a significant association between over-expression of telomerase enzyme and development of human cancers suggests new approaches to cancer therapy via combating this increased telomerase activity.

5. Repeated genomic sequences

Repeated sequences (also known as repetitive elements, or repeats) are patterns of nucleic acids (DNA or RNA) that occur in multiple copies throughout the genome. In many organisms, a significant fraction of the genomic DNA is highly repetitive, with over two-thirds of the sequence consisting of repetitive elements in human.
Repetitive elements found in genomes fall into different classes, depending on their mode of multiplication and/or structure. The disposition of repetitive elements consists either in arrays of tandemly repeated sequences, or in repeats dispersed throughout the genome.

These repeats represent potential source of genetic variation and regulation. Together with their regulatory roles, a structural role of repeated DNA in shaping the 3D folding of the genome has also been proposed.

There are 3 major categories of repeated genomic sequence:

1. Terminal repeats

2. Tandem repeats: copies which lie adjacent to each other, either in a direct or an inverted assembly.

a. Satellite DNA typically found in centromeres and heterochromatin.

b. Minisatellite repeat units from about 10 to 60 base pairs, found in many places in the genome, including the centromeres.

c. Microsatellite repeat units of less than 10 base pairs. They include the telomeres, which typically have 6 to 8 base pair repeat units.

3. Interspersed repeats

Transposable elements

SINEs (Short Interspersed Nuclear Elements)

LINEs (Long Interspersed Nuclear Elements)

SVAs

In primates, the majority of LINEs are LINE-1 and the majority of SINEs are Alu's. Alu elements are short stretches of DNA produced by the action of the restriction endonuclease enzyme Arthrobacter luteus (Alu). Alu elements are the most abundant transposable elements, containing over one million copies dispersed throughout the human genome. SVAs are hominoid specific.

In prokaryotes, CRISPR are arrays of alternating repeats and spacers.

Other types of genomic repeat sequences

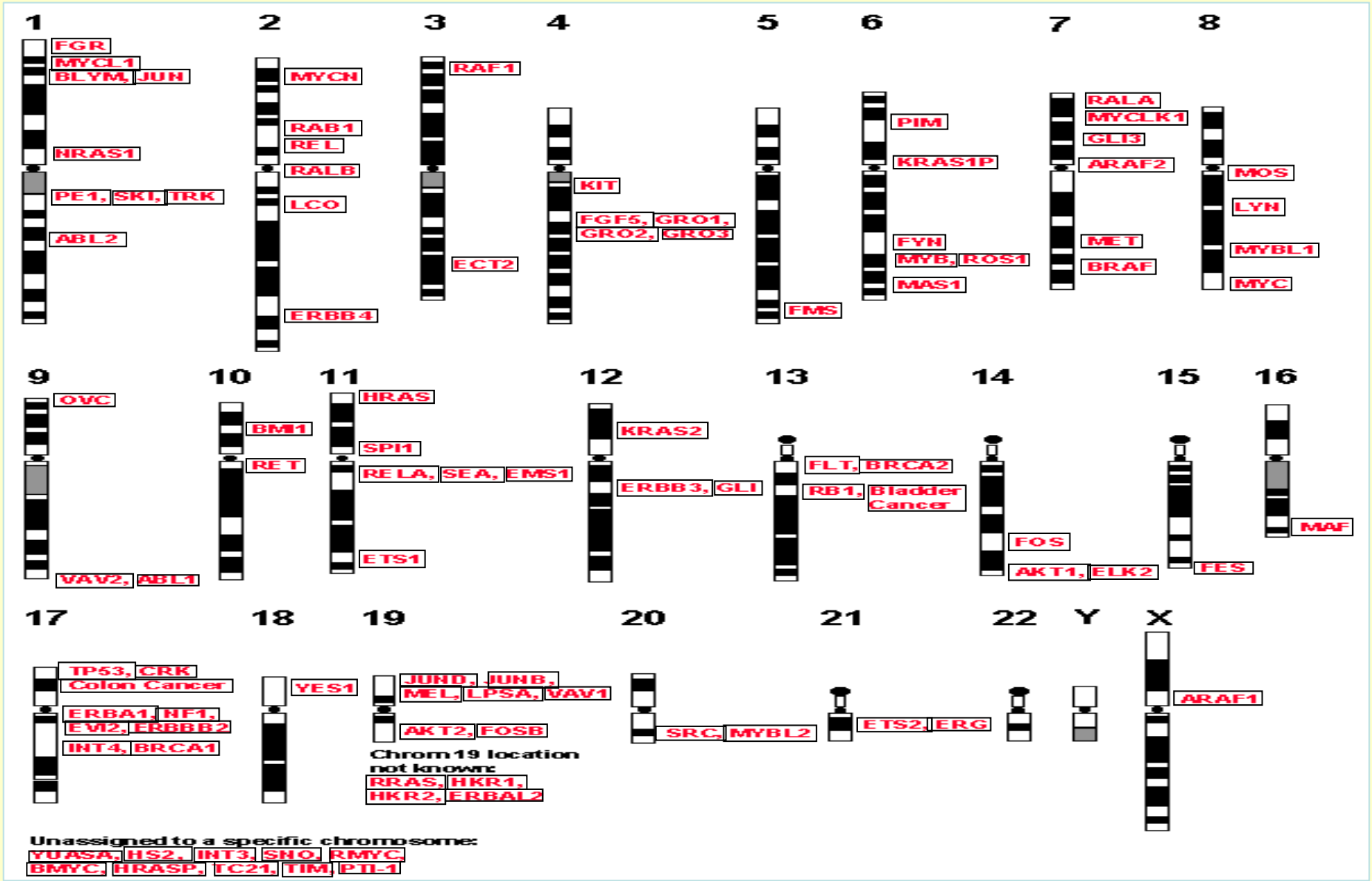
Direct repeats

Global direct repeat/Local direct simple repeats/Local direct repeats/Local direct repeats with spacer

Inverted repeats

Global inverted repeat/Local inverted repeat/Inverted repeat with spacer/Palindromic repeats/Mirror and everted repeats.

Distribution of Oncogenes, Cancer Genes and Tumor Suppressor Genes in the Human Genome

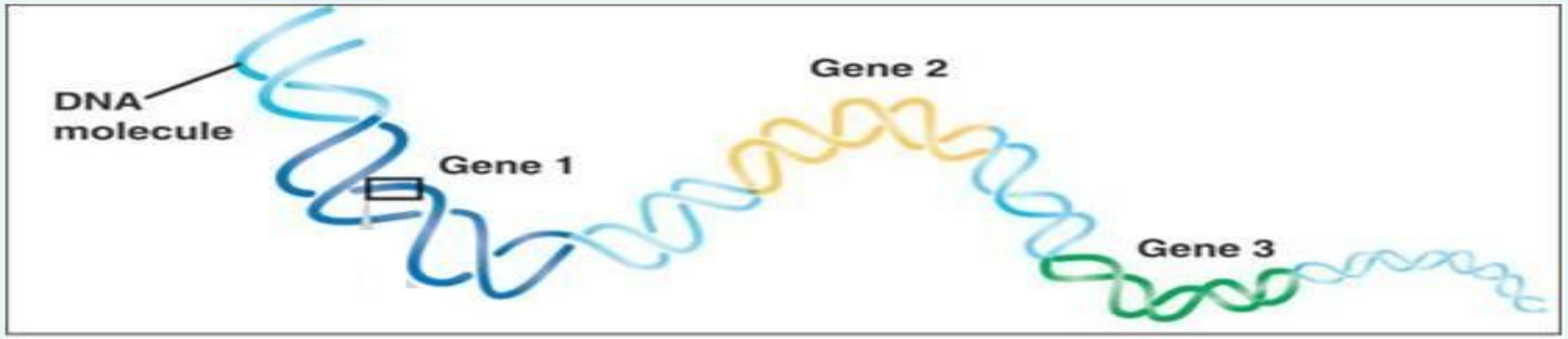


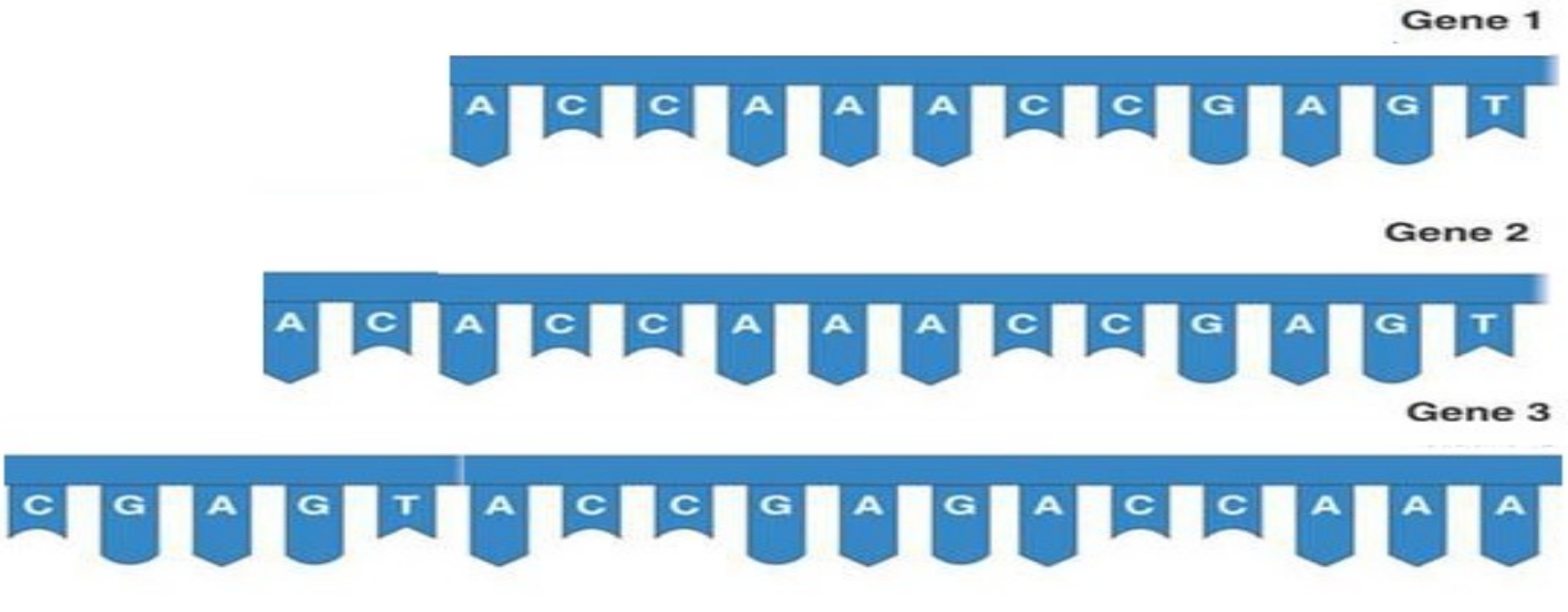
Structure Of Human Genes

The gene, which is the functional unit of the genome, is a specified linear sequence of nucleotides along one strand of DNA (the coding or active strand). The specific site of the gene on the DNA constituting a particular chromosome is called the gene locus and is characteristic of each gene. So, a gene might occupy a specific locus on the short arm or the long arm of the chromosome.

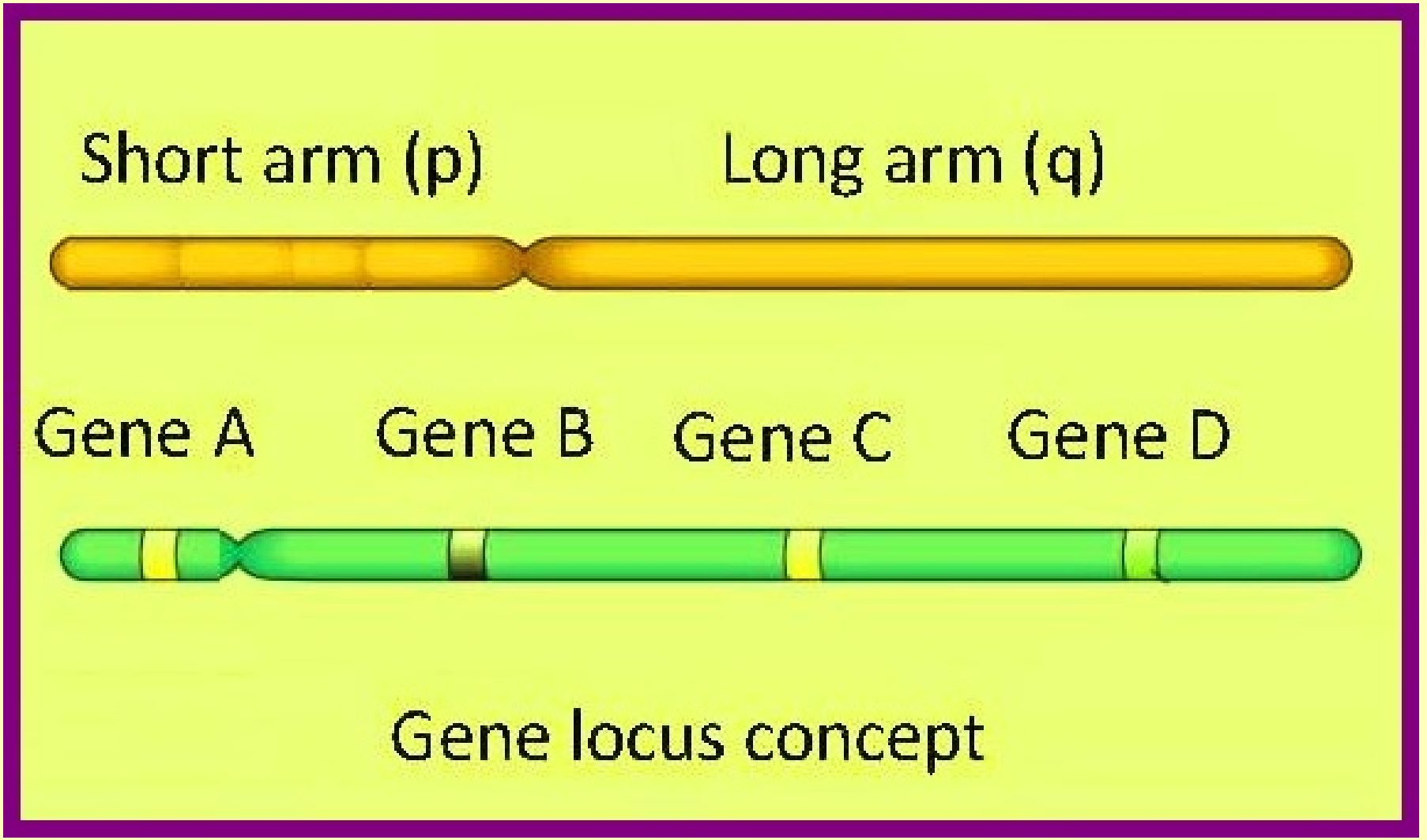
The gene occupies a specific site on one strand of DNA. If damage to the gene occurs, the other complementary strand is used as a template to repair the gene and replace defective parts of it by a specific DNA repair system.

Strucutre of Genes and Linear arrangement of genes along DNA

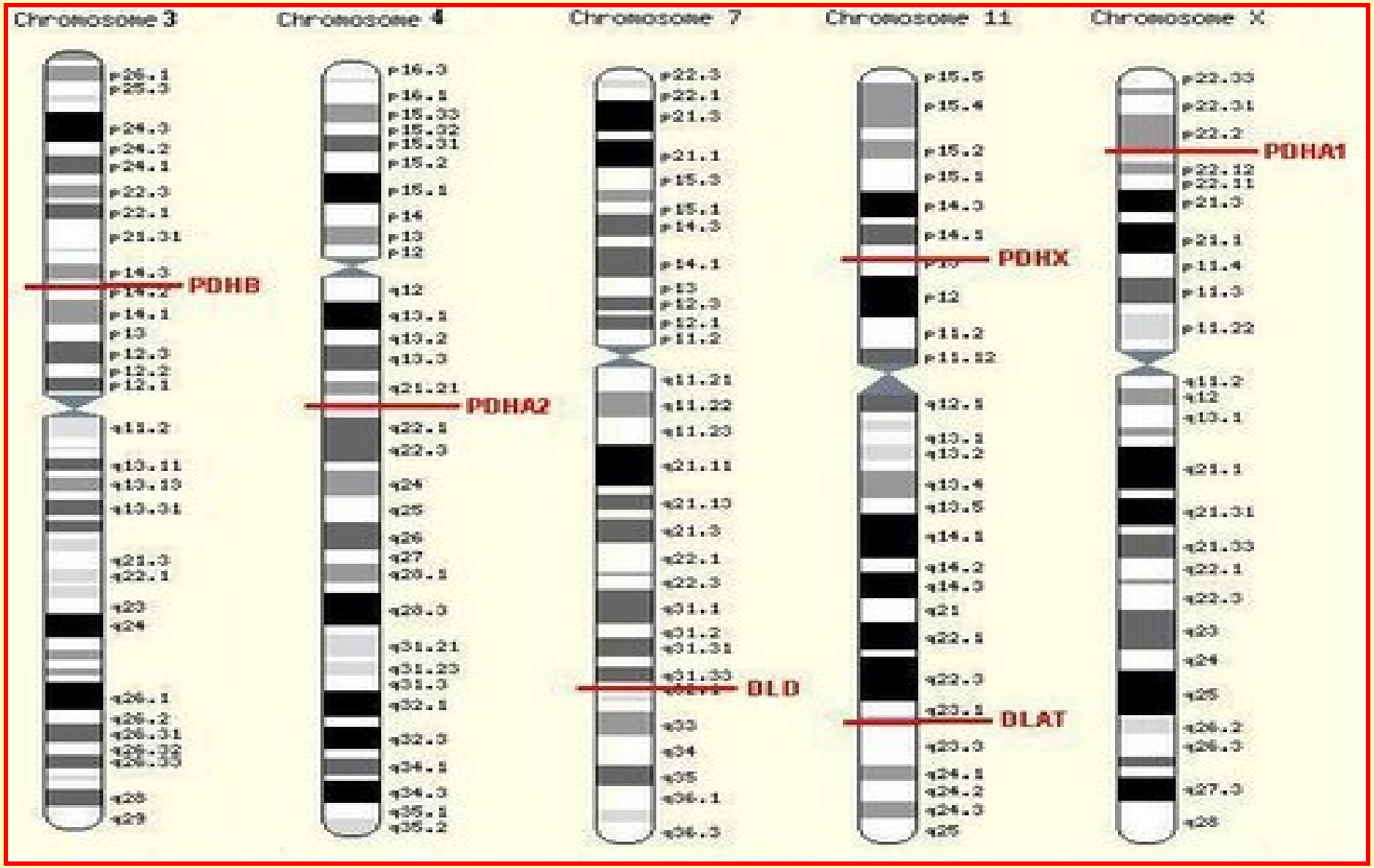




Linear Arrangement Of Genes Along DNA



The Concept Of Gene Locus



The estimated 38000 genes that comprise the nuclear genome constitute, and are distributed on, the 46 chromosomes in the nucleus. The longer and larger chromosomes have far more numbers of genes than the smaller and shorter chromosomes.

Genes are arranged in a linear sequence on chromosomes. Because genes constitute only a very small portion of the whole genome, they are separated by multiple long inter-genic parts of base sequences of the DNA that comprise most of the non-genic components of the genome. These include : transposons, transcriptional units, pseudo-genes, pyknons, long and short interspersed elements, among many other components of, yet, unknown function(s).

All genes have the same basic structure, being composed of a long piece of DNA consisting of the 4 nucleotides (A,G,C,T), but in varying numbers and peculiar arrangement characteristic of each gene.

Some genes are formed only of few hundred nucleotides, e.g. globin genes, while others consist of many hundred thousands up to 2.4 million nucleotides which constitute the Dystrophin gene.

The specific arrangement of the nucleotides of the gene imparts to each gene its functional specificity. Functionally, genes differ from each other by the structure and nature of the protein(s) synthesized under their regulation, which is determined by the specific arrangement of the nucleotides of the gene.

Types of Human genes

A. Structural Genes

1. Constitute the majority of genes.

2. Responsible for regulating synthesis of proteins through transcription and translation of mRNA).

B. Regulatory genes

Responsible for regulating functions of structural genes through transcription/silencing factors.

C. Master Genes

1. Responsible for maintaining the identity, the stability and the integrity of the genome/transcriptome/proteome.

2. Responsible for regulating higher vital cellular functions (production of ATP, DNA replication, cell division, apoptosis, transport across cell membranes, etc).

There are three main functionally-defined groups of genes in the human genome. These are : structural genes that are directly involved in protein synthesis, regulatory genes that control the function(s) of structural genes, and master genes which control and regulate the indespensable life activities of the cell including cell growth, cell division, cell differentiation, DNA repair, apoptosis and the like.

Structural genes regulate protein synthesis via the genetic information, or the genetic code, in the gene which is created and designed to function in a specific way so that each three nucleotides in sequence along the gene, known as the codon, which is the functional unit of the gene, define a single specific amino acid in the protein synthesized under regulation of the relevant gene.

So, according to the specific arrangement of the codons of a gene, a peculiar arrangement of amino acids in the protein synthesized by that gene occurs, leading to the synthesis of a specific protein thus imparting to each gene its functional specificity inspite of the common sharing of all genes in their basic building nucleotides.

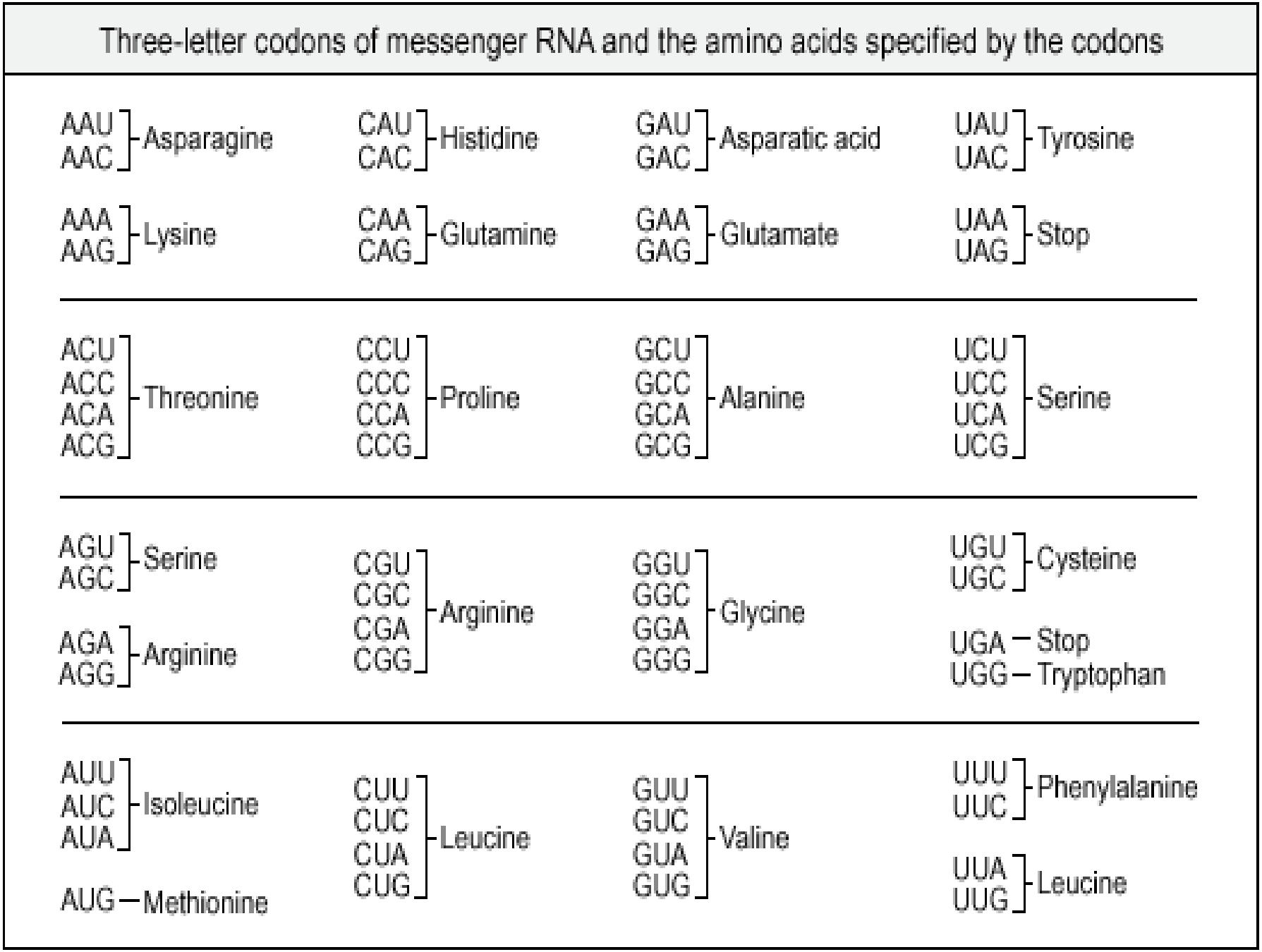
Thus, though all genes are formed of the same four nucleotides, an Insulin gene regulates the synthesis of Insulin, a Hemophilia gene regulates the synthesis of anti-hemophilic globulin, and a collagen gene regulates the synthesis of collagen depending on the specific amino acid sequence of each protein which is determined by the peculiar codon sequence of the relevant genes.

The Genetic Code

1. The gene is composed of nucleotides.
2. The protein is composed of amino acids.
3. The genetic code is the information embodied within the gene that allows it to define the synthesis of a particular protein based on the number and sequence of its bases.
4. The genetic code is designed so that each three bases in sequence (triplet or codon) define a specific amino acid in the synthesized protein.
5. As the nucleotide is the structural unit of the gene, the codon is the functional unit of the gene. There are 64 codons, 61 active codon specifying amino acids and three (3) stop or termination codons that do not specify any amino acids.

The Genetic Code

TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys
TTC	Phe	TCC	Ser	TAC	Tyr	TGC	Cys
TTA	Leu	TCA	Ser	TAA	STOP	TGA	STOP
TTG	Leu	TCG	Ser	TAG	STOP	TGG	Trp
CTT	Leu	CCT	Pro	CAT	His	CGT	Arg
CTC	Leu	CCC	Pro	CAC	His	CGC	Arg
CTA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CTG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser
ATC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
ATA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
ATG	Met*	ACG	Thr	AAG	Lys	AGG	Arg
GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly
GTC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GTA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GTG	Val	GCG	Ala	GAG	Glu	GGG	Gly


Three-letter codons of messenger RNA and the amino acids specified by the codons
AAU
AAC
Asparagine
CAU
CAC
Histidine
GAU
GAC
Asparatic acid
UAU
UAC
Tyrosine
AAA
AAG
Lysine
CAA
CAG
Glutamine
GAA
GAG
Glutamate
UAA
UAG
Stop
ACU
ACC
ACA
ACG
Threonine
CCU
CCC
CCA
CCG
Proline
GCU
GCC
GCA
GCG
Alanine
UCU
UCC
UCA
UCG
Serine
AGU
AGC
Serine
AGA
AGG
Arginine
CGU
CGC
CGA
CGG
Arginine
GGU
GGC
GGA
GGG
Glycine
UGU
UGC
Cysteine
UGA — Stop
UGG — Tryptophan
AUU
AUC
AUA
Isoleucine
AUG — Methionine
CUU
CUC
CUA
CUG
Leucine
GUU
GUC
GUA
GUG
Valine
UUU
UUC
Phenylalanine
UUA
UUG
Leucine

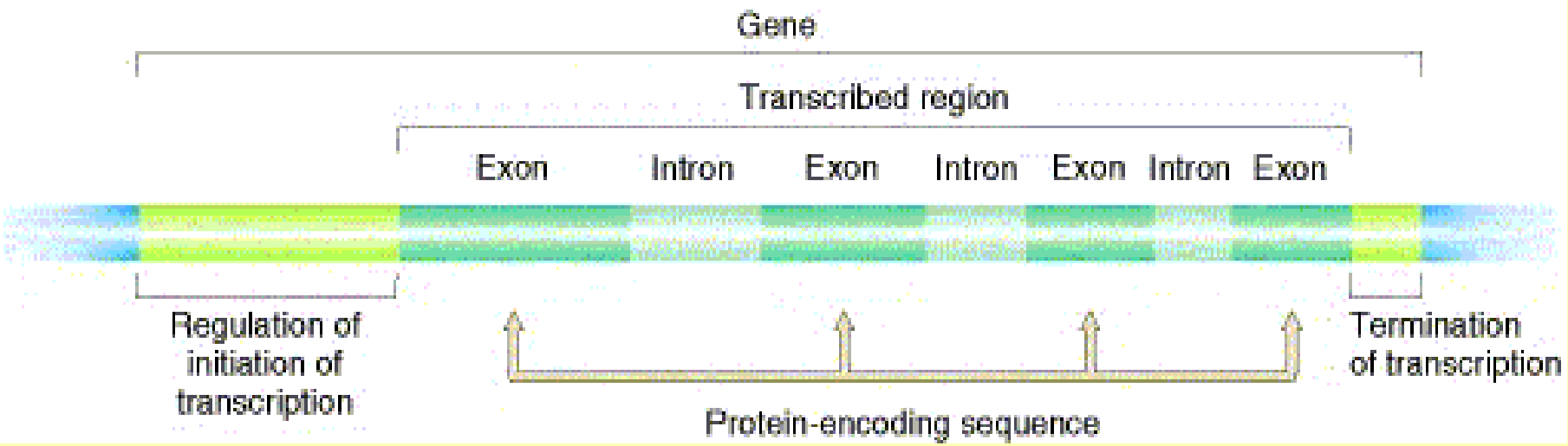
Stages Of Gene Function

1. Gene switching (stimulation/activation).
2. Transcription (synthesis of mRNA).
3. Post-transcription modifications of m-RNA

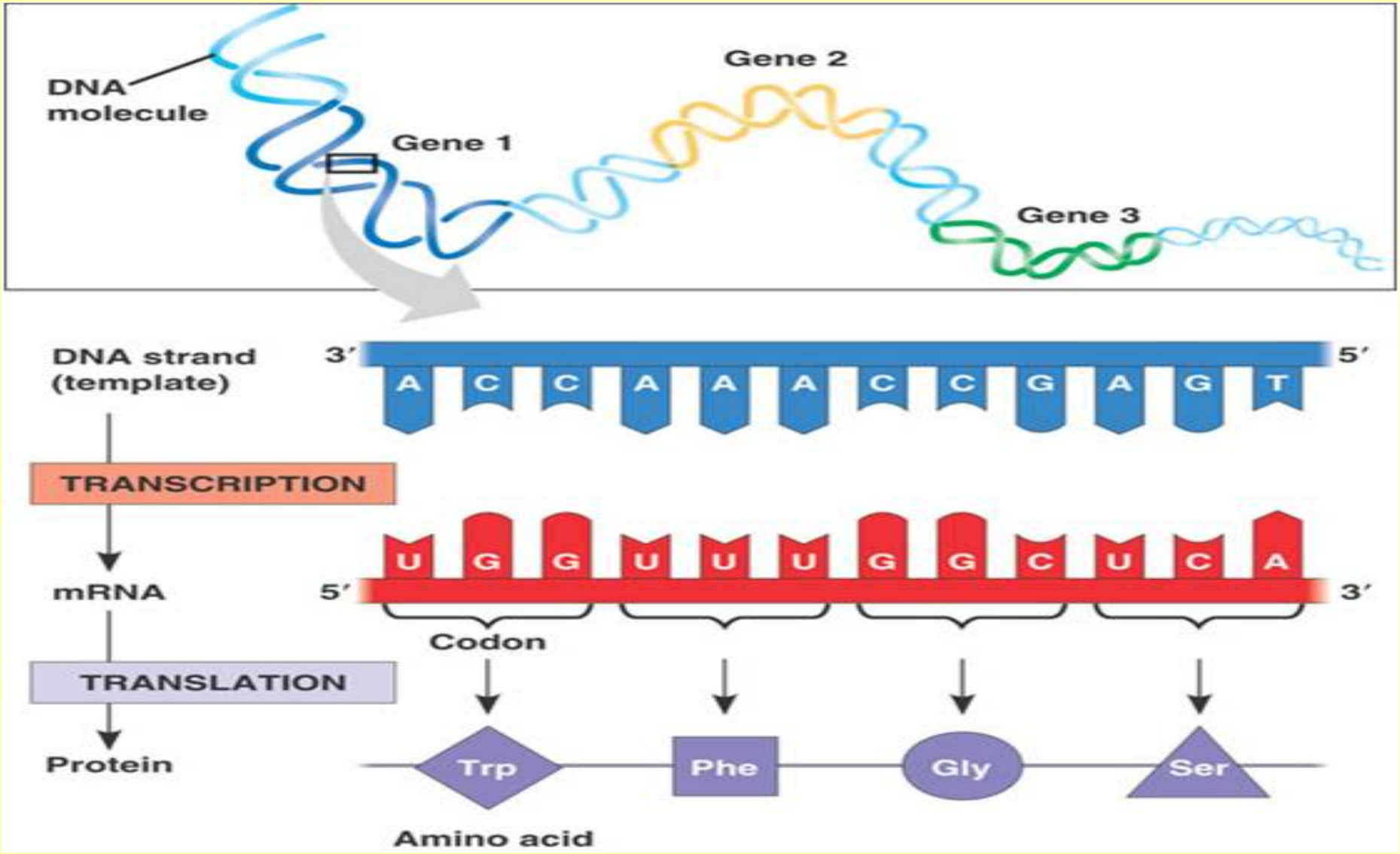
   (removal of introns and splicing of exons, addition of poly adenylate tail and many other changes)
4. Translation (synthesis of protein).
5. Post-translation modifications of Proteins

   (folding, addition of other components, etc).
6. Post-translation Trafficking of Proteins.

Functionally, the gene consists of three main parts : the promotor area responsible for switching on the gene to start function or switching it off to stop function, the exons which are the parts of the gene responsible for defining the amino acids of the protein synthesized by the gene, and the introns which are the parts of the gene which, with some exceptions, do not participate in protein synthesis.



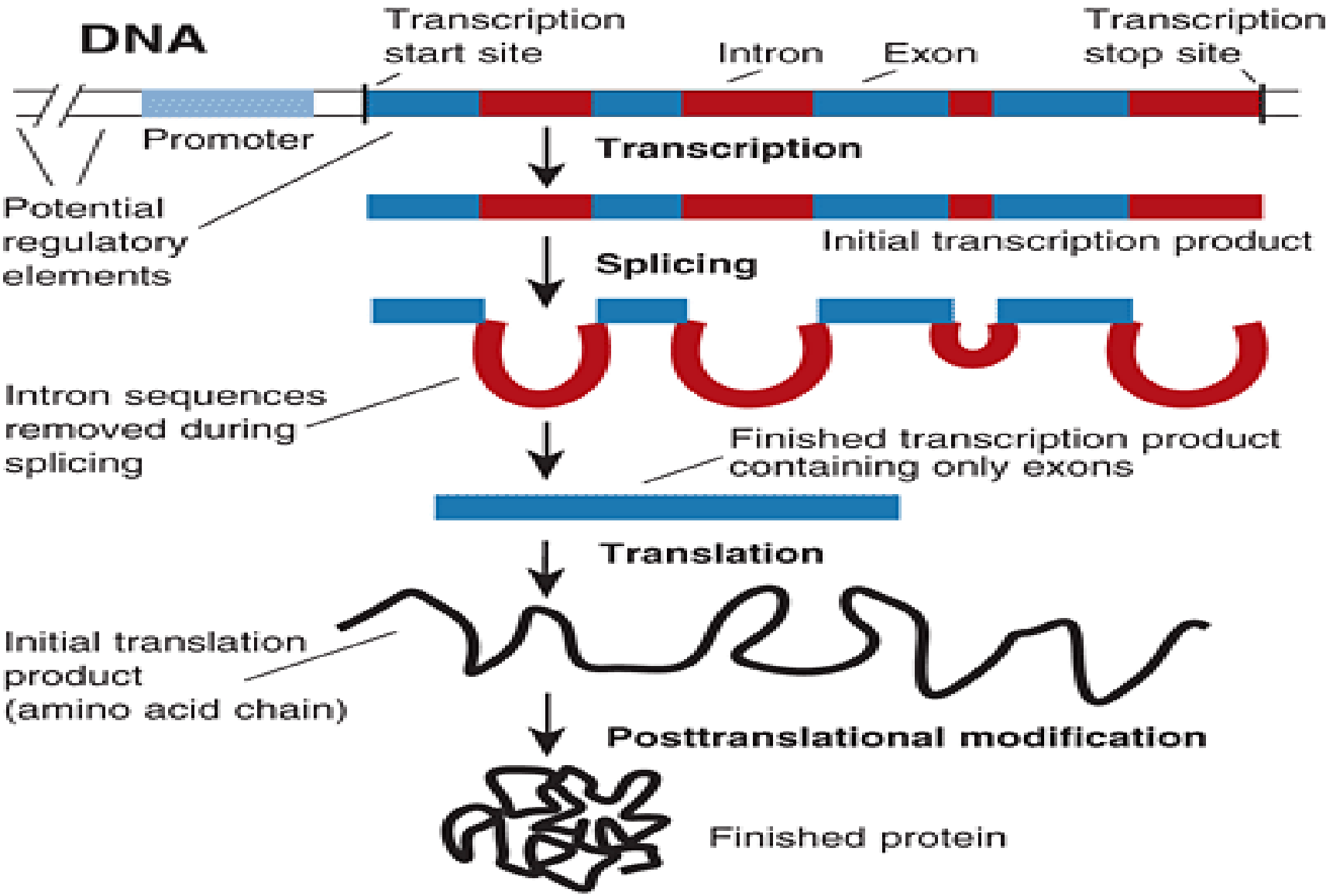
Stages Of Gene Function



The Genetic Code

TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys
TTC	Phe	TCC	Ser	TAC	Tyr	TGC	Cys
TTA	Leu	TCA	Ser	TAA	STOP	TGA	STOP
TTG	Leu	TCG	Ser	TAG	STOP	TGG	Trp
CTT	Leu	CCT	Pro	CAT	His	CGT	Arg
CTC	Leu	CCC	Pro	CAC	His	CGC	Arg
CTA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CTG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser
ATC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
ATA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
ATG	Met*	ACG	Thr	AAG	Lys	AGG	Arg
GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly
GTC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GTA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GTG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Stages Of Gene Function

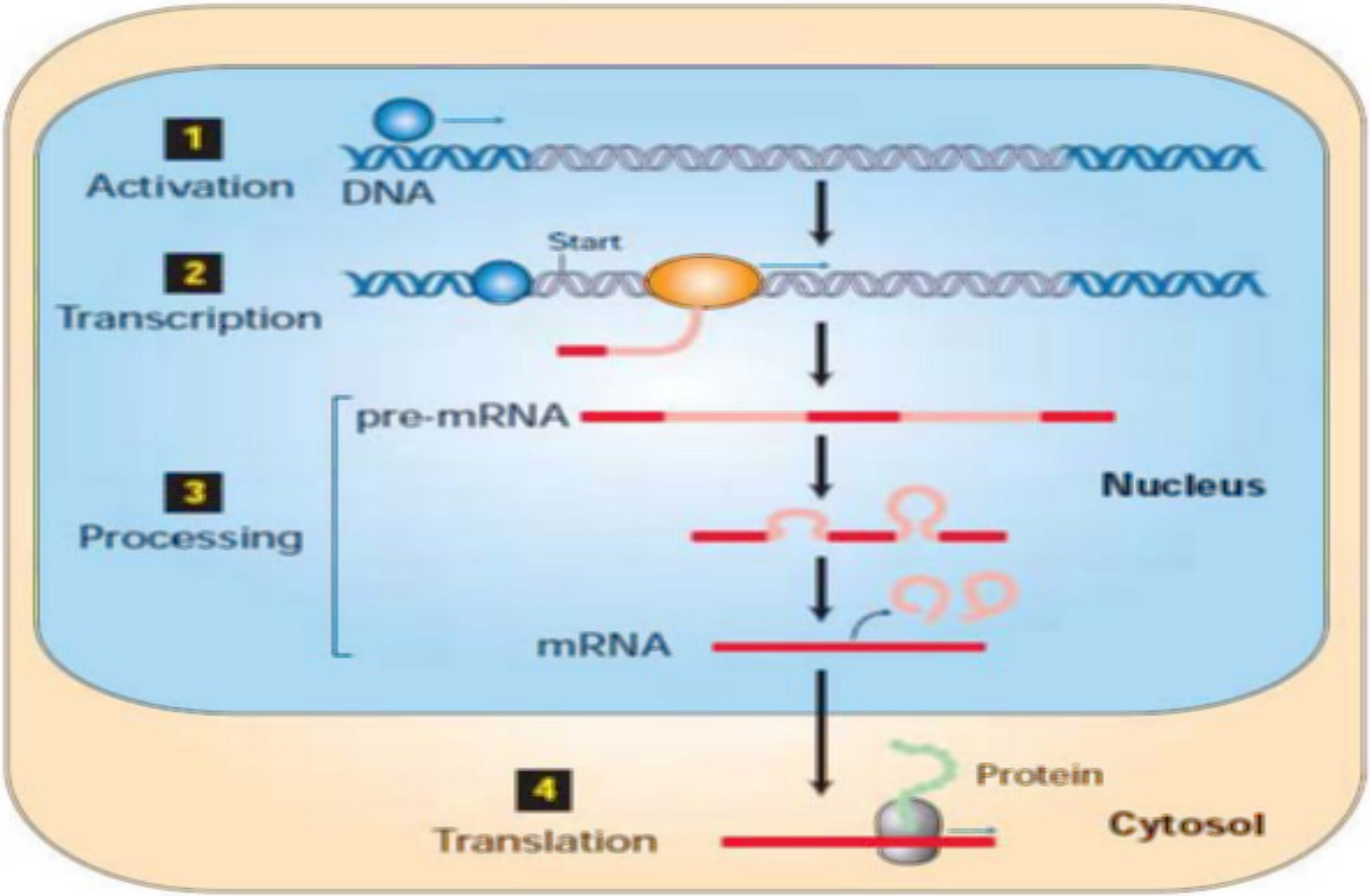


Introns are alternatively distributed along the gene with the exons and are removed from the mRNA in a process involving excision or removal of introns and splicing or joining of the exons. This process is a pre-requisite for synthesis of proper active proteins, otherwise, larger, unstable, easily degradable, physiologically non-functioning proteins, might be synthesized.

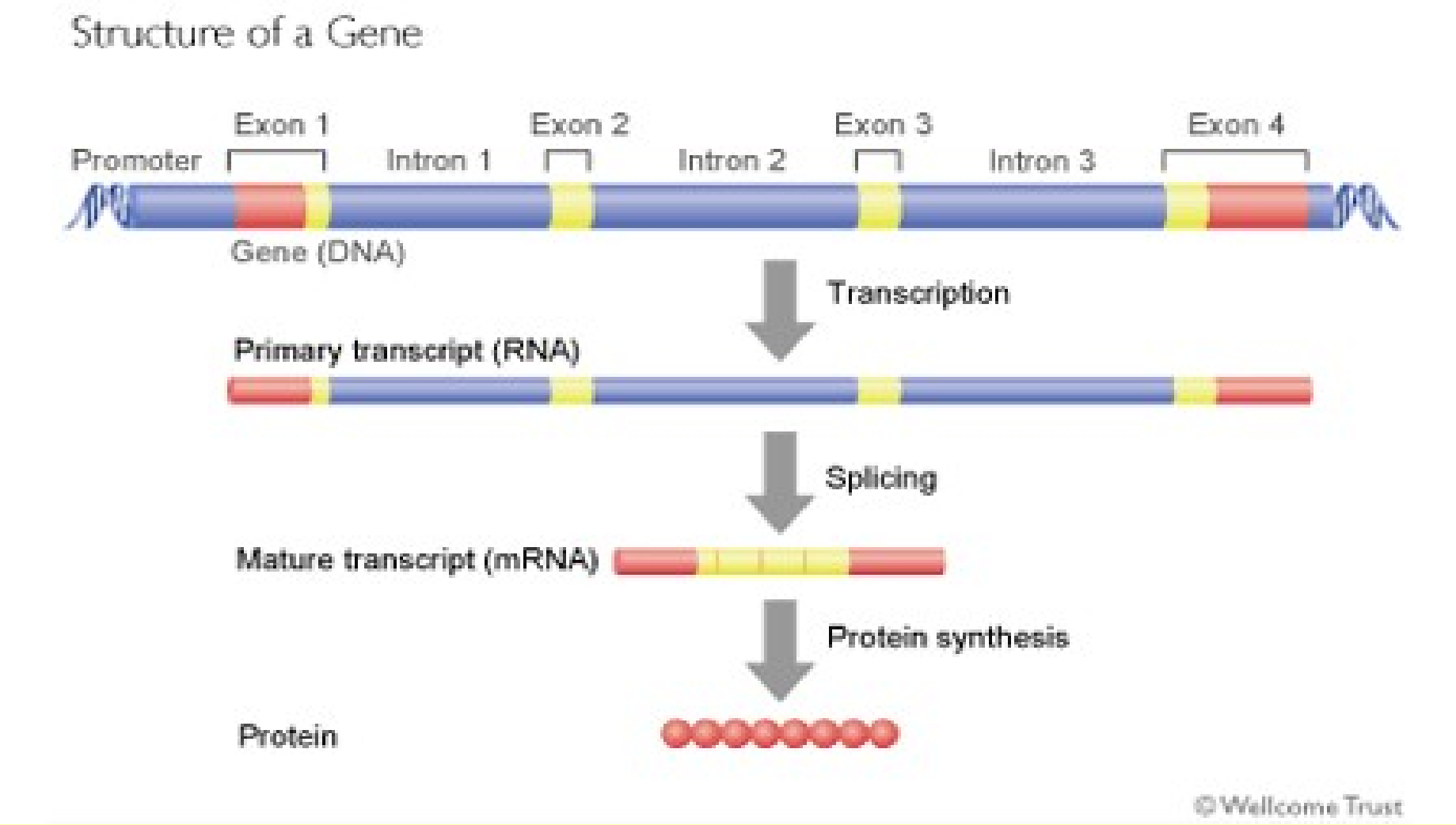
However, in some genes introns are kept in the mRNA and are translated into the protein, and via alternative removal of one or more of these introns, the gene can code for the synthesis of more than one protein.

This feature of alternative intron excision explains the huge number of proteins (nearly 400000 – 4000000) that constitute the human proteome produced under control of the far less number of genes (nearly 25000 – 38000) that constitute the human genome.

Excision of Introns and Splicing of Exons



Excision of Introns and Splicing of Exons



Stages Of Protein synthesis

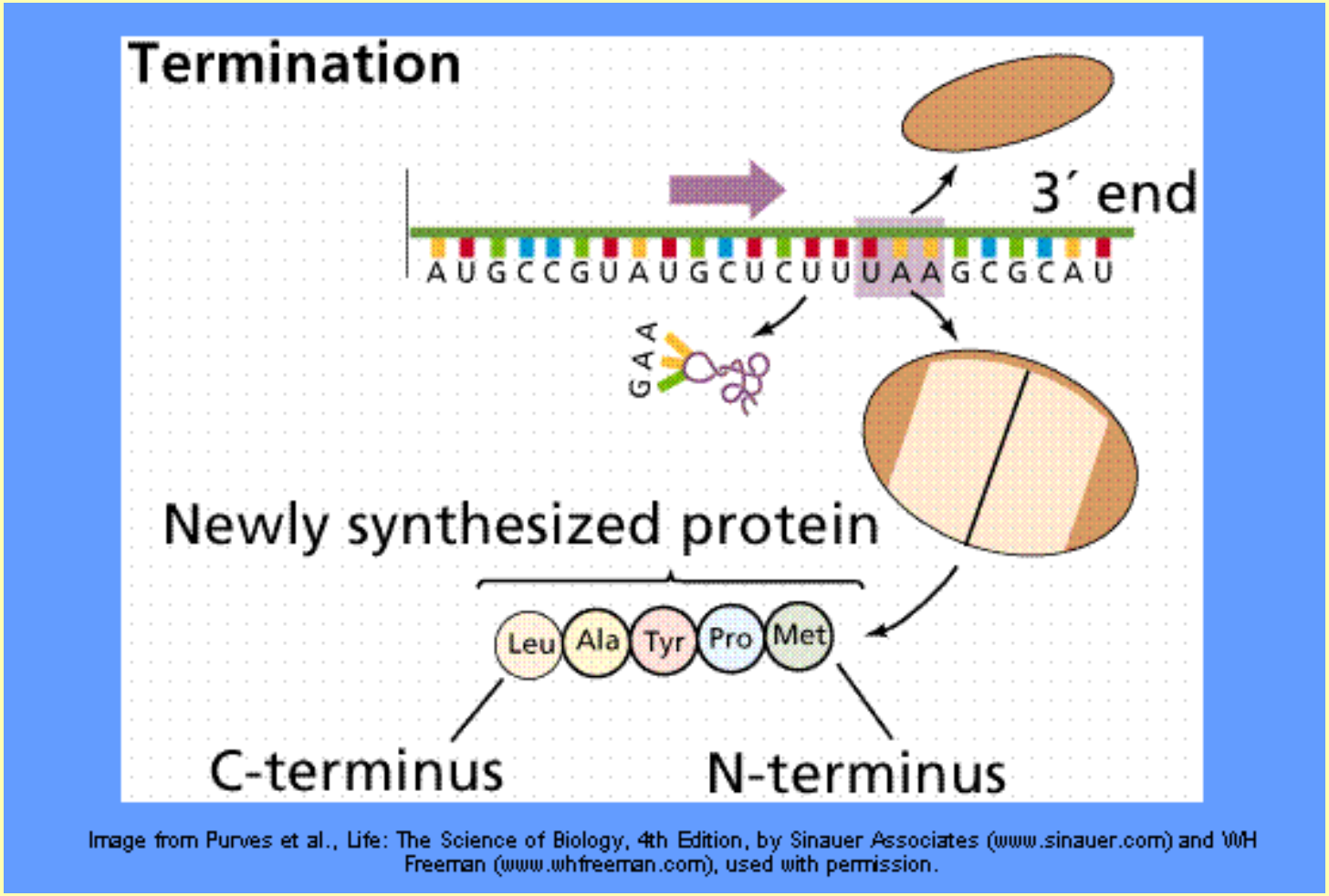
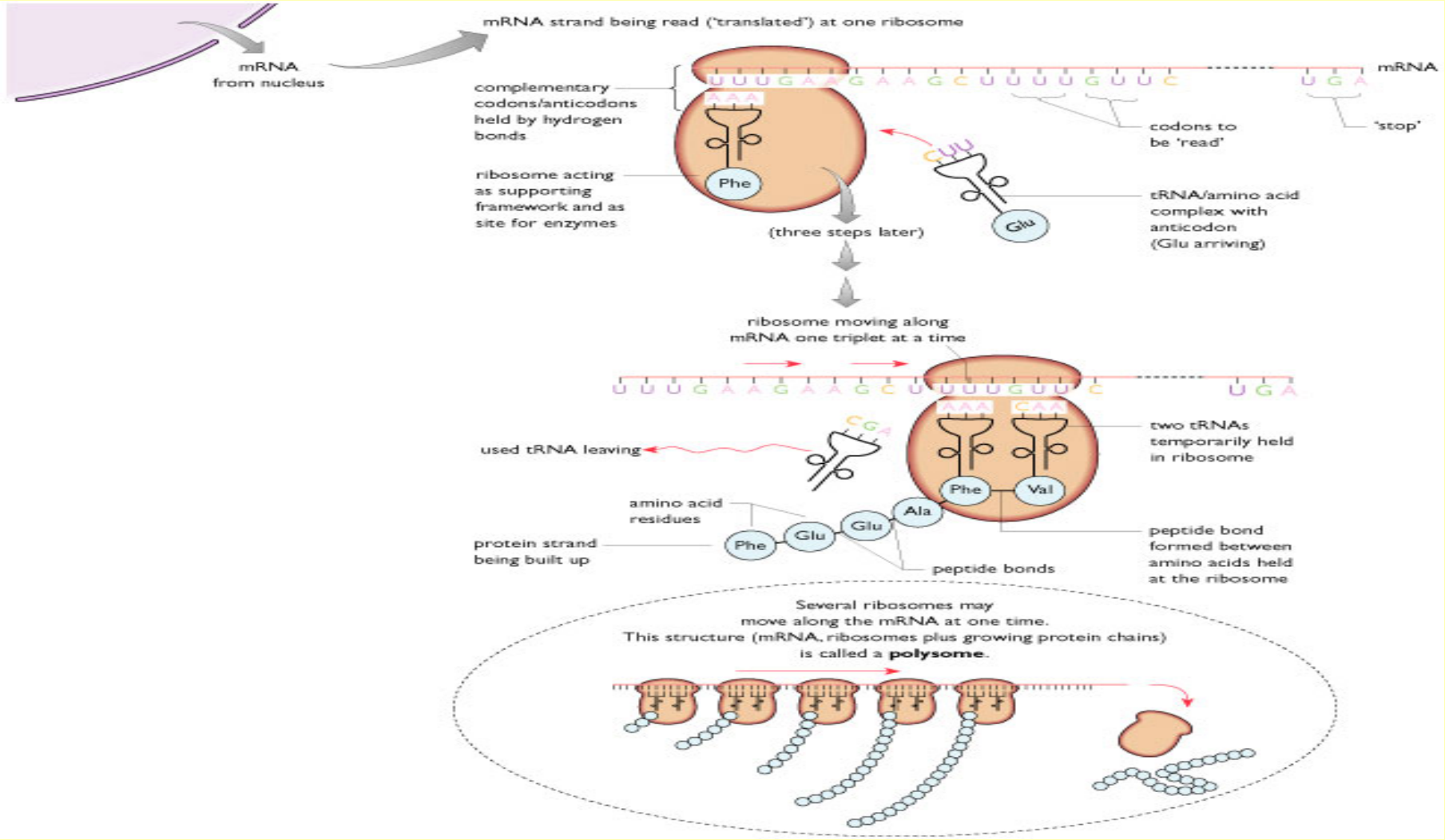


Image from Purves et al., Life: The Science of Biology, 4th Edition, by Sinauer Associates (www.sinauer.com) and WH Freeman (www.whfreeman.com), used with permission.

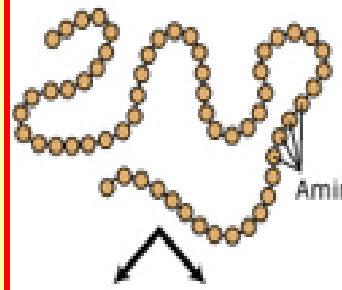
Translation : From Codons to Amino acids
Synthesis of Proteins in The Cytoplasm



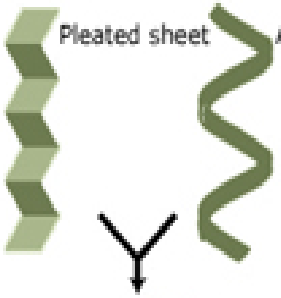
Post-translation Modifications of Proteins

The majority of newly synthesized proteins must undergo specific structural modifications, e.g. folding, to become functionally active. These modifications of protein structure are very important and critical for most proteins to confer upon them physiological potency.

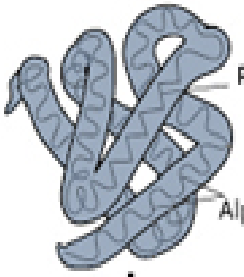
Failure of completing these structural modifications leads to production of defective proteins and underlies the development of a large number of serious genetic diseases like alph-1 antitrypsin deficiency and many immunodeficiency disorders.



Primary protein structure is sequence of a chain of amino acids.



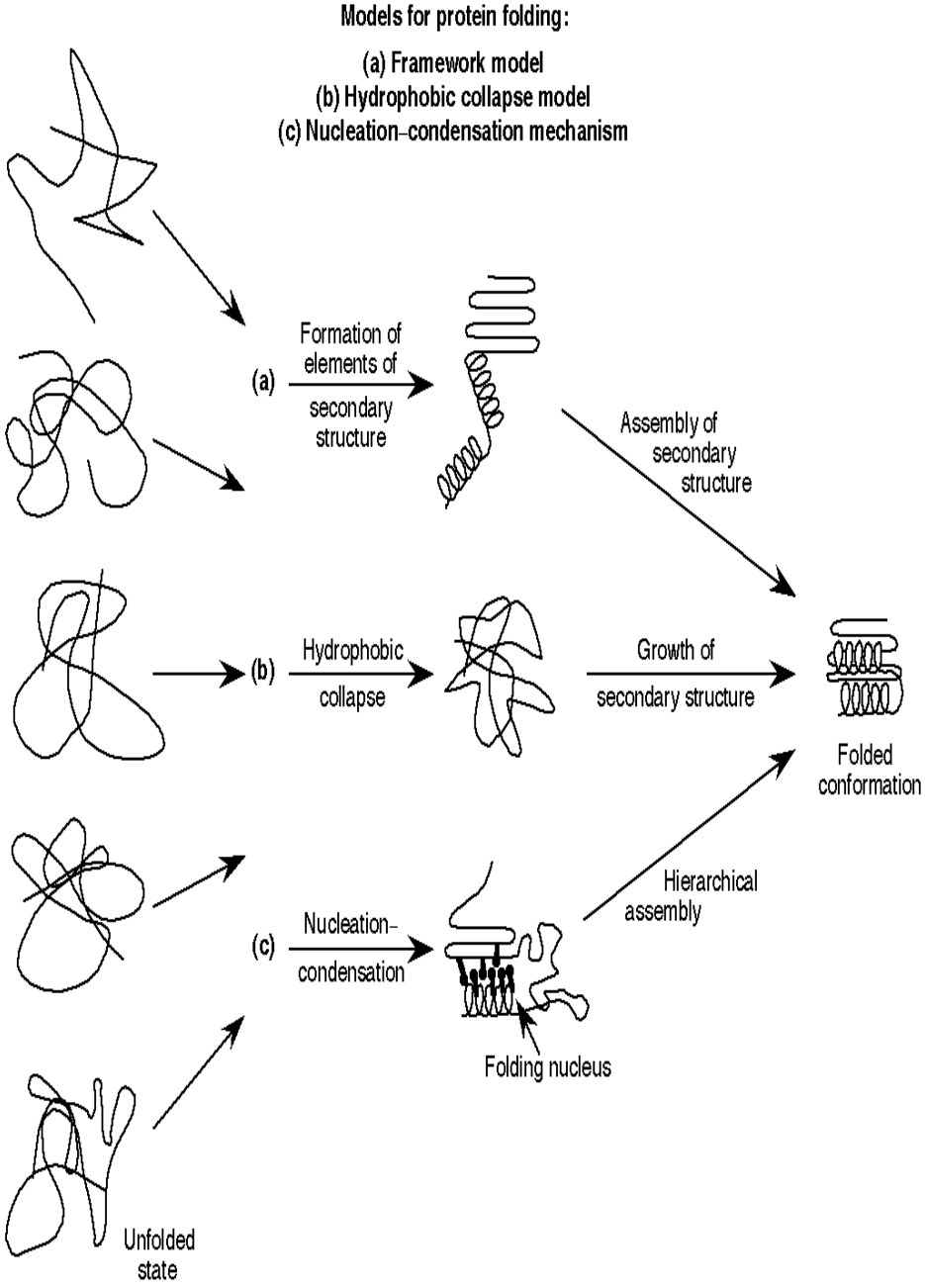
Secondary protein structure occurs when the sequence of aminoacids are linked by hydrogen bonds.



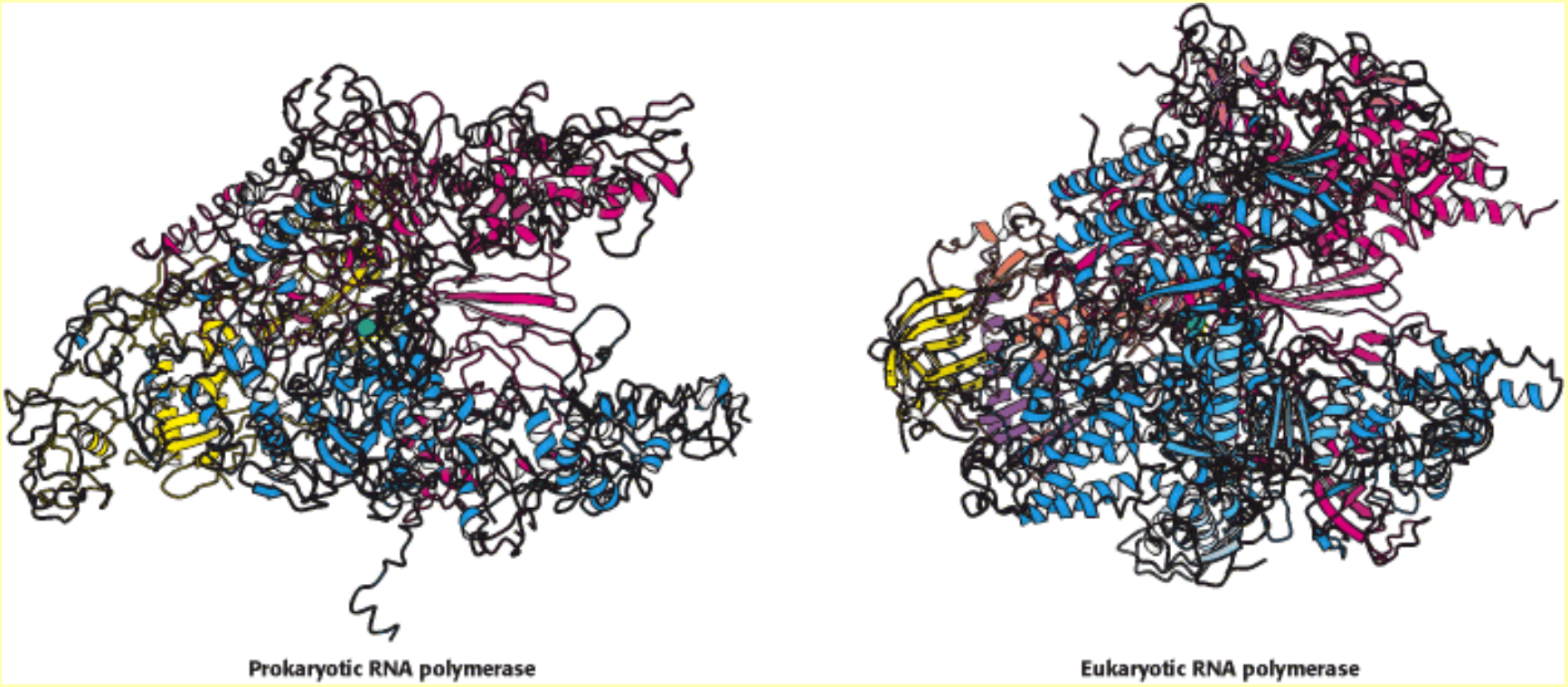
Tertiary protein structure occurs when certain attractions are present between alpha helices and pleated sheets.

Quaternary protein structure is a protein consisting of more than one amino acid chain.

Image adapted from: National Human Genome Research Institute.



Structure Of RNA Polymerase after undergoing post-translational modifications



Post-translation trafficking of proteins

Life activities within living cells are mediated by proteins. Two major classes of proteins can be recognized within this functional context: structural proteins, like the cytoskeleton and cell membrane proteins, and catalytic proteins or enzymes that actually conduct and regulate all metabolic networks encompassing biological processes within the cell. Proteins are highly specialized biomolecules. Their functional specialization is intimately dependent on their proper localization at their targeted sites of action inside the cell. Hence, newly synthesized proteins within the cytoplasm of the cell have to be transported, trafficked or targeted, from their site of

synthesis and directed, or targeted, to their sites of action within the cell, e.g. insertion in cellular membranes, cell organelles or catalysis of metabolic activities inside the mitochondria.

Post-translation trafficking of proteins refers to the dynamic processes that follow synthesis of new proteins, aiming at their proper localization within the cell compartments.

Trafficking is critical for proper functioning of proteins. Precise targeting of proteins depend on synthesis of specific factors, mostly short amino acid sequences or chemical

molecules like mannose-6-phosphate, that direct the transport of the protein to its exact destination.

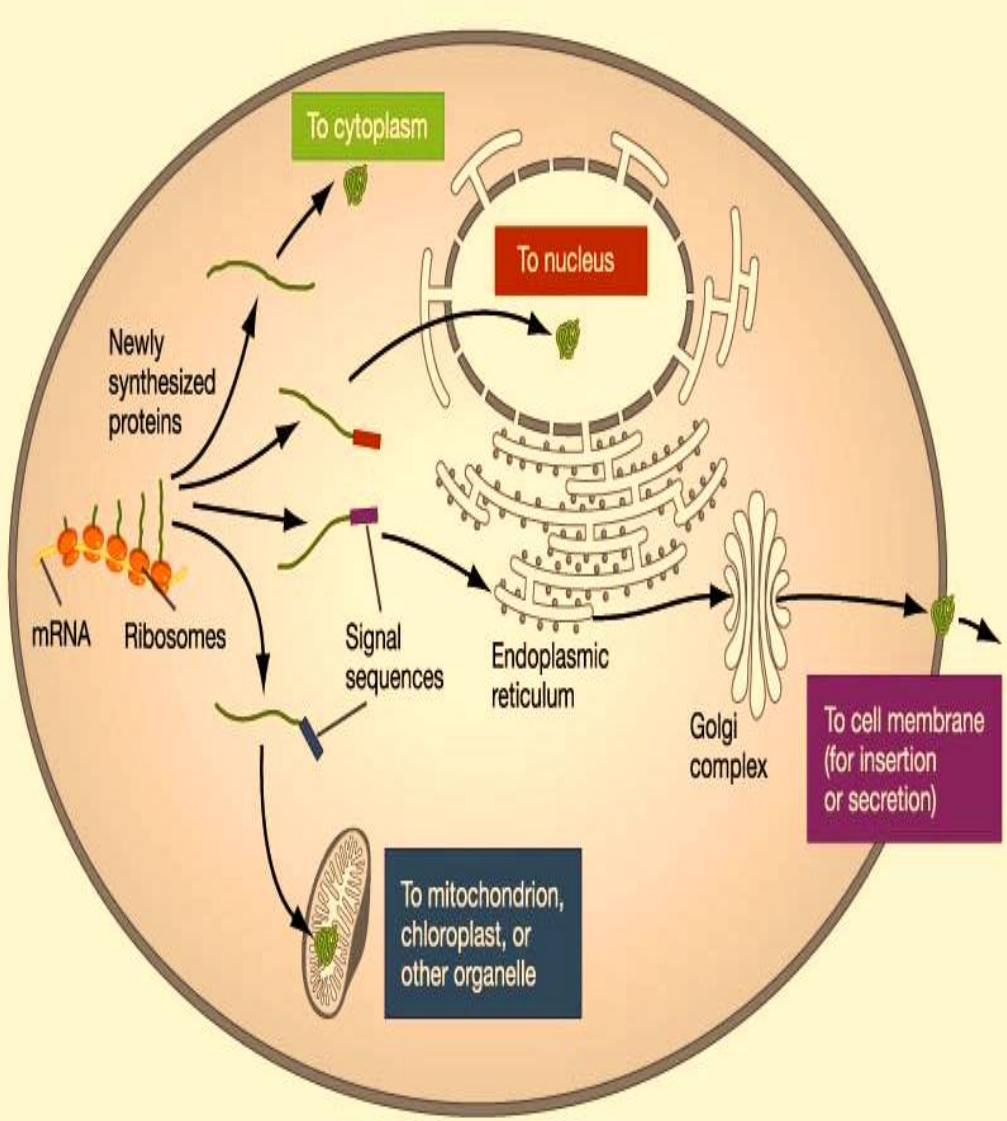
Post-translation trafficking of proteins involves active participation of the endoplasmic reticulum and the Golgi apparatus. Passage through one or both of these organelles is necessary for many proteins to become active biomolecules or to get ready for attachment to their specific recognition signal molecules needed for trafficking to their proper sites.

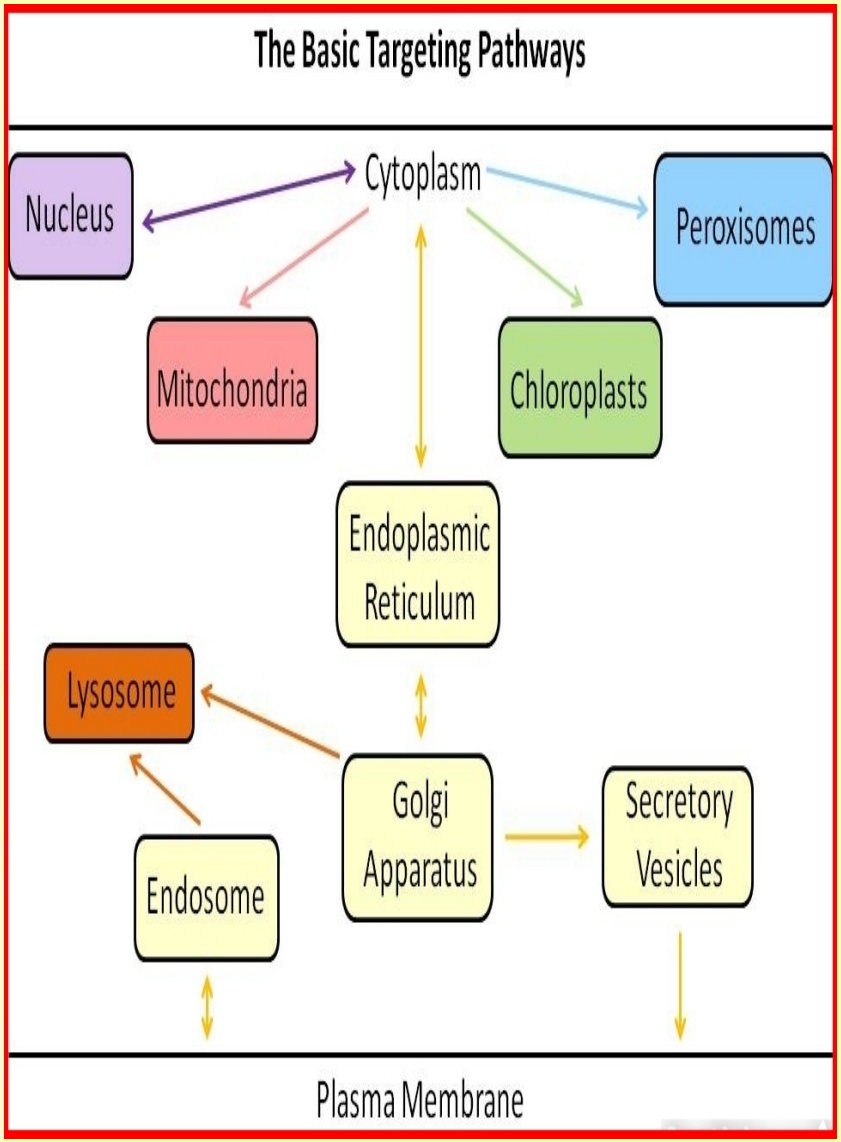
The rough endoplasmic reticulum is an integral part of the protein targeting pathway. Proteins that pass

through it and exit from there are marked with signal sequences that work as address label directing the proteins to their destination. In the absence of protein targeting signals, newly formed proteins remain functionless at their sites of synthesis in the cytoplasm.

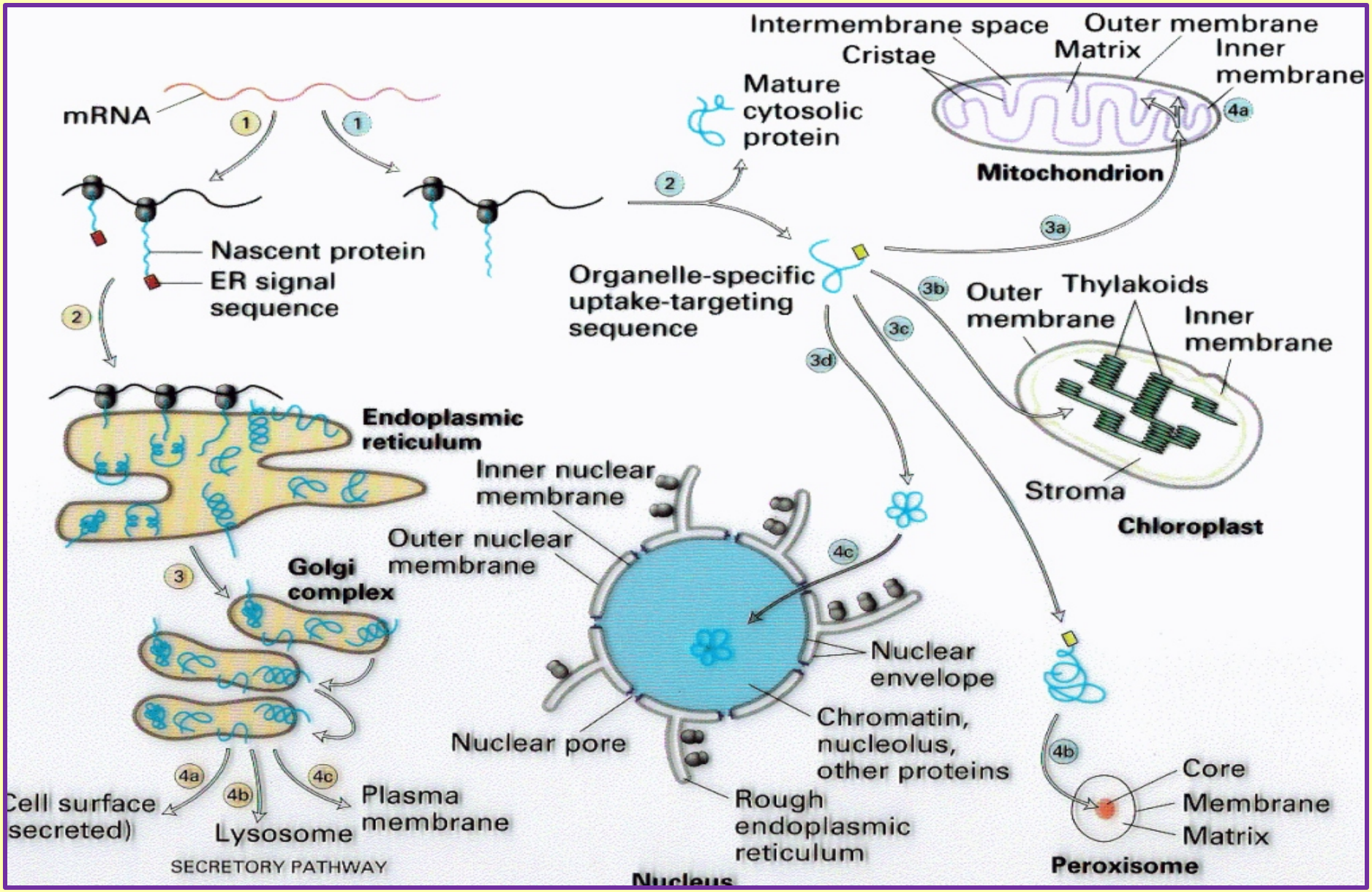
Defects in trafficking processes can result in disturbed localization of the protein to its target site with resultant functional deficiency of its biological and/or metabolic activities. Many genetic diseases are caused by failure of targeting properly synthesized proteins from their sites of synthesis to their sites of action.

Post-translation trafficking of proteins

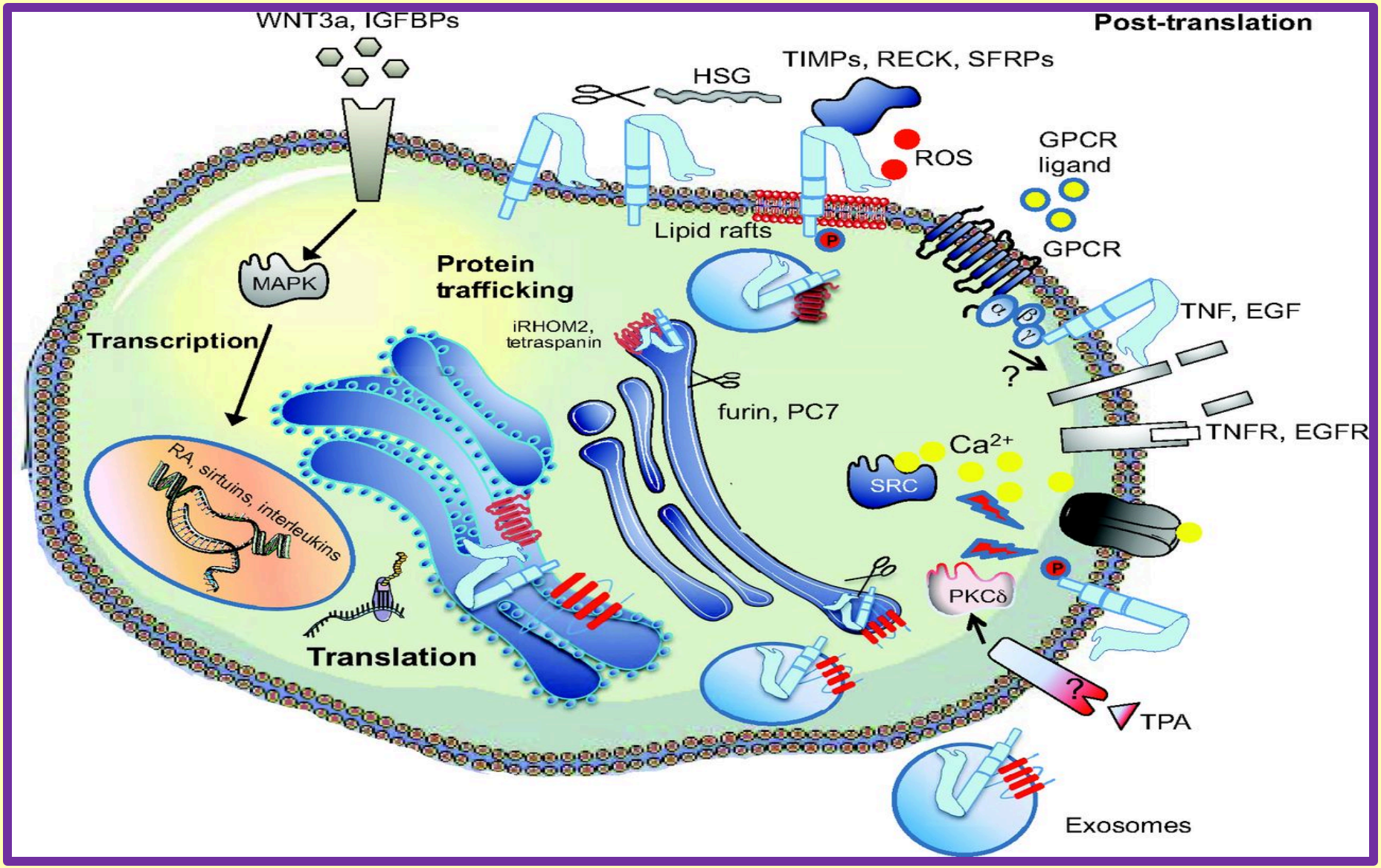




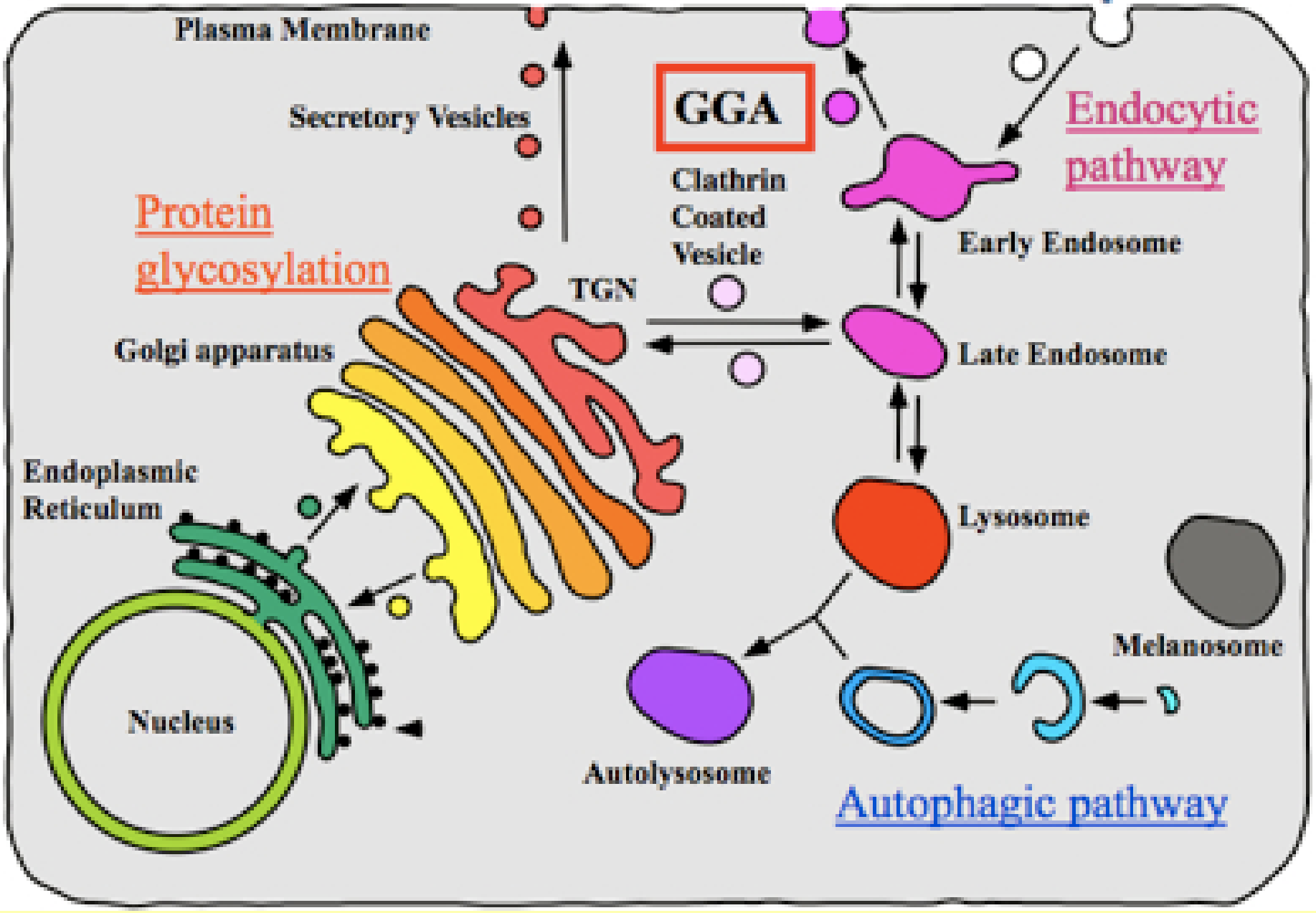
Protein Trafficking



Protein Trafficking & Transport



Post-translational modification and transport



www.archive.org/details/MedicalGenetics